PROJECT BEFORE YOU DISTILL: CARRYING PRIVILEGED CHAIN-OF-THOUGHT ACROSS SENSOR GAP

Anonymous authors
Paper under double-blind review

ABSTRACT

Large vision–language models provide rich reasoning priors, but physical sensors often lack large paired corpora and intermediate reasoning supervision. Directly distilling a rationale generated from a privileged sensor can introduce claims that are not supported by the deployment sensor. To address this gap, this paper introduces *Cross-sensor Rationale Projection* (CRP) which treats the privileged rationale as a reasoning proposal and re-derives it against the deployment observation. CRP keeps supported information, abstracts fine details into sensor-supported structure, hedges partially supported claims, and removes unsupported content, followed by a consistency screen. We further analyze information loss across the sensor gap and derive an *Estimated Projection Value* (EPV), computable even before training the deployment student model, to indicate when projection helps with consistent empirical support. Two purpose-built sensor reasoning benchmarks, RadarQA-300 and AudioQA-300, are constructed for RGB-to-radar and audio-visual-to-audio reasoning. CRP consistently outperforms answer-only supervision, direct privileged-rationale distillation, and deployment-sensor rationale supervision, with gains of 5.9 and 12.9 percentage points over answer-only training on RadarQA-300 and AudioQA-300, respectively.

1 INTRODUCTION

Large vision–language models are typically trained on extensive image–text corpora, thereby acquiring strong priors for visual representation learning and multimodal reasoning (Bai et al., 2025; Shao et al., 2024a). In contrast, physical sensing systems, including millimeter-wave radar, audio, depth, thermal, and inertial signals (An et al., 2022; Fan et al., 2026; Li et al., 2022), often yield non-RGB observations. Relative to image–text data, these modalities are supported by markedly fewer paired datasets and provide limited chain-of-thought or rationale-style supervision. Consequently, existing sensor–language models tend to remain narrowly tailored to specific downstream applications (e.g., sensor captioning or task-specific perception) rather than general-purpose reasoning (Lai et al., 2026; Han et al., 2026). The central challenge, therefore, extends beyond label acquisition to the development of intermediate reasoning supervision that is both faithful to and compatible with the sensor measurements encountered at deployment.

Learning with privileged information (e.g., RGB, depth), followed by cross-modal distillation constitutes a natural starting point, wherein an information-rich sensor stream is leveraged during training to provide supervisory signals for a model that, at inference time, operates solely on the target deployment sensor modality (Vapnik & Vashist, 2009; Lopez-Paz et al., 2016; Huo et al., 2024; Zhao et al., 2026). Vision–language models can further transform privileged observations into intermediate rationales (Shao et al., 2024a; Xu et al., 2025a; Dong et al., 2025). A straightforward strategy is therefore to distill such privileged rationales to fine-tune the deployment model. However, rationales grounded in the privileged sensor may encode information that is not supported by, or recoverable from, the deployment sensor observation. In the radar setting, for instance, this may include appearance- or identity-specific details that the deployment model is then implicitly expected to reconstruct from radar measurements alone, which is infeasible (see Figure 1). An alternative is to forgo rationale-level supervision and instead supervise only the final prediction via answer-only supervised fine-tuning. While this mitigates the risk of imposing unsupported intermediate targets,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

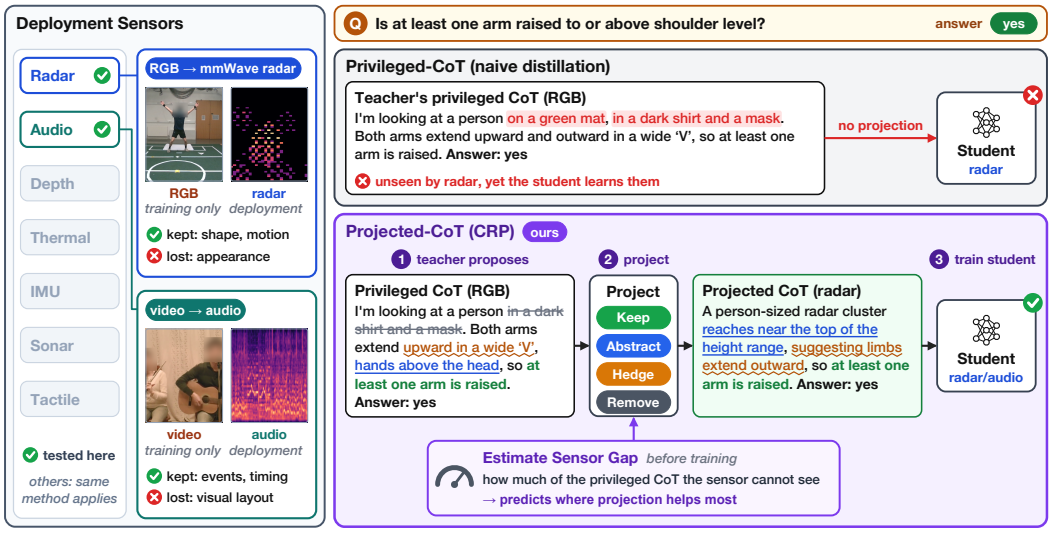

Figure 1: Overview of CRP. *Left.* We study two privileged-to-deployment sensor settings: RGB→mmWave radar and video→audio, illustrating information retained and lost across the sensor gap. *Right.* Privileged-CoT directly transfers the RGB rationale and can therefore supervise the radar student with claims not supported by the deployment sensor. In contrast, CRP treats the privileged rationale as a proposal, projects its claims using the deployment sensor, and trains the student. The sensor gap is estimated before training to indicate where projection is beneficial.

it provides limited direct guidance regarding the inference process. A further alternative is to employ a strong multimodal teacher to generate rationales directly from the deployment modality. This yields sensor-native reasoning, but it places on the teacher the burden of jointly selecting useful intermediate steps and ensuring that each step is justifiable given only the deployment observation.

In this paper, we introduce *Cross-sensor Rationale Projection* (CRP), depicted in Fig. 1. CRP treats a rationale generated from the privileged observation as a reasoning proposal and re-derives (i.e., *projects*) it against the deployment sensor before using it as supervision. During projection, a multimodal teacher rewrites the privileged reasoning using evidence supported by the deployment sensor. CRP keeps supported information, abstracts fine details into sensor-observable structure, hedges partially supported claims when relevant distinctions remain unresolved, and removes unsupported content. A subsequent consistency screen checks the projected rationale against the deployment observation and the final answer before it is used to train a student that receives only the deployment observation. The privileged and deployment observations therefore play complementary roles: the privileged observation proposes intermediate reasoning that may be difficult to recover from the deployment sensor alone, while the deployment observation determines which parts of that reasoning can be supported and how they should be expressed.

Our main contributions are:

- We introduce *Cross-sensor Rationale Projection* (CRP), a framework for transferring intermediate reasoning from a privileged observation to a deployment sensor. Rather than directly distilling the privileged rationale, CRP re-derives its claims against the deployment observation through keep, abstract, hedge, and remove operations, followed by a consistency screen before training deployment sensor models.
- We develop an analysis of information loss across the sensor gap under a simplified abstract model. Using the privileged rationale as a common reference, we distinguish the answer-relevant reasoning lost when a rationale is generated directly from the deployment sensor from the loss that remains after projection.
- We derive an *Estimated Projection Value* (EPV) from paired privileged–deployment sensor observations and rationale supervision before student training. Empirical results confirm that EPV is a useful indicator of where projection provides greater benefit.

- We construct two purpose-built sensor reasoning benchmarks, RadarQA-300 and AudioQA-300, covering RGB-to-radar and audio-visual-to-audio reasoning. Projected-CoT achieves the highest answer accuracy across models, fine-tuning schemes, and sensor representation formats.

2 Related Work

Rationale supervision. Chain-of-thought supervision provides intermediate reasoning signals beyond final answers (Wei et al., 2023), and has been extended to vision–language models through visual rationales and structured reasoning (Xu et al., 2025a; Dong et al., 2025). However, directly transferring teacher rationales can introduce unsupported or imitative reasoning (Wu et al., 2025; Chen et al., 2025). Recent methods therefore reconstruct, attribute, or restrict privileged reasoning to evidence recoverable from the student observation (Tian et al., 2026; Zhang et al., 2026). Our setting considers a stronger sensor gap: the privileged and deployment observations come from physically different sensors, and CRP re-evaluates the claims of the privileged rationale against the deployment observation before using them for supervision.

Cross-modal distillation. Cross-modal methods transfer information between sensing modalities. C2KD performs bidirectional cross-modal distillation and filters misaligned soft labels (Huo et al., 2024), while PULSE aligns deployable-sensor features with a privileged teacher through a shared representation (Zhao et al., 2026). These approaches primarily transfer predictions or representations. In contrast, CRP transfers a natural-language reasoning chain and re-evaluates its individual claims against the deployment sensor.

Reasoning with radar and audio. RadarLLM aligns radar point-cloud sequences with language for human-motion understanding without explicit chain-of-thought reasoning (Lai et al., 2026), while RAGent uses synchronized video to provide semantic knowledge for radar reasoning (Han et al., 2026). RAGent is particularly related in its use of privileged visual information, but transfers activity semantics through retrieval rather than projecting a privileged reasoning chain through the radar observation. For audio, MUSIC-AVQA provides audio-only, visual-only, and audio-visual question answering over musical performances (Li et al., 2022; Liu et al., 2024). We use these two settings to evaluate the same rationale-projection principle across distinct sensor gaps.

3 Problem Formulation

We consider reasoning tasks in which richer sensing is available during training than at deployment. Each training instance contains a privileged sensor x^p (e.g., RGB, depth), a deployment sensor x^d (e.g., radar or audio), a natural-language question q, and a ground-truth answer $\bar{y}$. A teacher model can access the privileged observation to construct reasoning supervision, whereas the student model receives only (x^d, q) and has no access to x^p at inference time. The goal is to transfer useful reasoning from the privileged sensing setting to the deployment student.

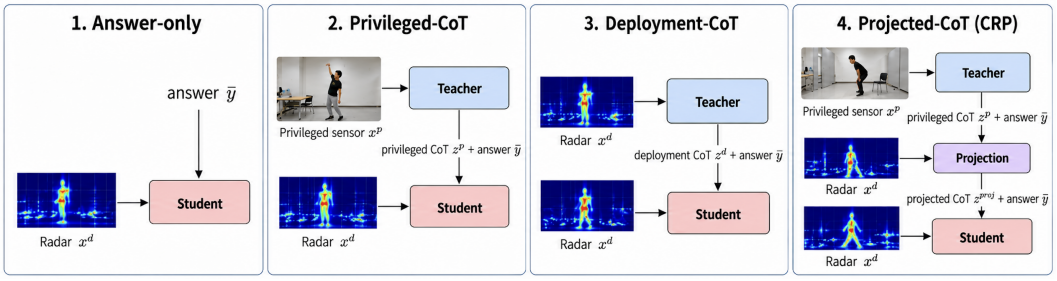

Figure 2: **Cross-modal supervision strategies.** Answer-only uses only the final answer, Privileged-CoT distills reasoning from the privileged sensor, Deployment-CoT generates reasoning directly from the deployment sensor, and Projected-CoT (CRP) projects the privileged rationale into reasoning supported by the deployment sensor.

The difficulty is the sensor gap: the privileged observation may reveal answer-relevant evidence that is missing or not supported in the deployment sensor, making a rationale valid under x^p potentially unsuitable for x^d. Figure 2 illustrates three direct supervision strategies. The simplest is *Answer-only*, where the student is trained only on the ground-truth answer $\bar{y}$. This avoids rationale errors but discards intermediate reasoning supervision. A second strategy is *Privileged-CoT*: a teacher model generates a rationale $z^p = T_p(x^p, q, \bar{y})$ from the privileged observation, and the student is trained on $(z^p, \bar{y})$ while still receiving only (x^d, q) for training. A third strategy is *Deployment-CoT*. A teacher model instead generates a rationale directly from the deployment observation $z^d = T_d(x^d, q, \bar{y})$ and the student is trained on $(z^d, \bar{y})$. Privileged-CoT can retain rich reasoning but introduce claims unsupported by the deployment sensor, whereas Deployment-CoT respects the deployment observation but can lose useful reasoning available from the privileged view. Answer-only avoids both forms of rationale error by providing no intermediate reasoning supervision.

4 Cross-sensor Rationale Projection

CRP introduces an alternative, *Projected-CoT*, which starts from the privileged rationale and re-derives it against the deployment observation before using it as supervision. Our design is guided by two considerations: 1) retaining useful reasoning available from the privileged observation, and 2) ensuring that the resulting rationale is supported by the deployment observation.

Propose. The proposal stage combines a privileged sensor observation with a capable multimodal teacher. The privileged observation x^p provides richer task-relevant information than the deployment sensor observation, while the teacher provides the reasoning capability needed to organize that information into useful intermediate steps. Formally, a strong teacher T_p generates

$$z^p = T_p(x^p, q, \bar{y}), \tag{1}$$

from the privileged observation x^p, question q, and ground-truth answer $\bar{y}$. We use z^p as a proposal of *what to reason about*, rather than as supervision that is assumed to be directly valid for the deployment sensor. The proposal may identify useful entities, relations, and intermediate deductions that are difficult to discover from x^d alone.

We first decompose the privileged rationale into atomic claims,

$$\mathcal{C}^p = (c_1^p, \ldots, c_m^p) = \mathcal{D}(z^p), \tag{2}$$

where each c_i^p is an observational, relational, or derived statement whose support can be assessed against an observation or preceding reasoning. The decomposition is semantic rather than sentence-based. For example, the privileged rationale

> *"The left hand is above the shoulder while the right arm remains lower, so at least one arm is raised."*

may be decomposed as c_1^p = the left hand is above the shoulder, c_2^p = the right arm remains lower, c_3^p = at least one arm is raised.

Project: The projection stage determines how the information in the privileged rationale should be expressed under the deployment observation. The sensor-conditioned projection operator maps the privileged claims to projected claims $\tilde{c}_i$,

$$\widetilde{\mathcal{C}} = (\tilde{c}_1, \ldots, \tilde{c}_m) = \mathcal{P}_d\left(\mathcal{C}^p; x^d, q, \bar{y}\right), \tag{3}$$

where x^d constrains what information can be retained, and q and $\bar{y}$ guide the projection toward answer-relevant reasoning. Specifically, it associates privileged claims with one of four projection types $a_i \in \{\textsc{Keep}, \textsc{Abstract}, \textsc{Hedge}, \textsc{Remove}\}$ with the corresponding projected claim $\tilde{c}_i$:

- **Keep.** The privileged claim is supported by the deployment observation at its original granularity and is retained without changing its semantic content:

$$a_i = \textsc{Keep} \quad \Longrightarrow \quad \tilde{c}_i \equiv c_i^p. \tag{4}$$

  Examples include coarse spatial position or temporal ordering that is directly observable from the deployment sensor.

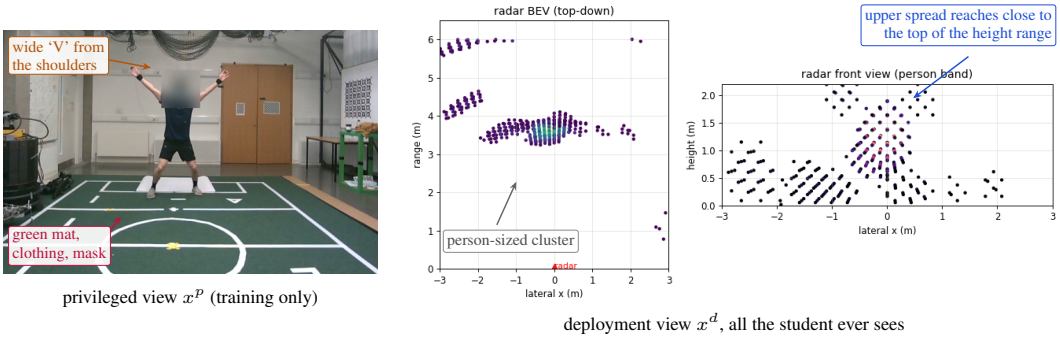

Figure 3: **Radar rationale projection on one training example.** The privileged rationale z^p is generated from RGB, while the projected rationale z^{proj} is re-derived from the paired radar observation x^d and then screened. Colors indicate claim-level projection operations: keep, abstract, hedge, and remove. The example shows how appearance-specific details are removed or weakened while radar-supported evidence for the final answer is retained.

- **Abstract.** The privileged claim contains finer distinctions than the deployment observation supports, but a coarser sensor-supported statement can be retained:

$$a_i = \textsc{Abstract} \implies \tilde{c}_i \prec c_i^p, \tag{5}$$

  where $\tilde{c}_i \prec c_i^p$ indicates that $\tilde{c}_i$ preserves a less specific version of the information in c_i^p. For example, a detailed visual description of body configuration may be rewritten using the coarser upper-body structure supported by radar.

- **Hedge.** The deployment observation supports part of the privileged claim but leaves a relevant distinction unresolved. The projected claim retains the supported information while explicitly marking the unresolved component:

$$a_i = \textsc{Hedge} \implies \tilde{c}_i \prec c_i^p \quad \text{with unresolved information made explicit.} \tag{6}$$

  For example, "the left arm is raised" may become "an arm appears raised, but its side cannot be resolved from radar."

- **Remove.** The privileged claim cannot be supported by the deployment observation, or is contradicted by it, and no projected claim is retained:

$$a_i = \textsc{Remove} \implies \tilde{c}_i = \varnothing. \tag{7}$$

Typical radar examples include clothing color, facial appearance, and other fine visual attributes absent from the radar observation.

More examples of both radar and audio claim operations a_i are shown in Appendix D.

The projection types are selected jointly over $\mathcal{C}^p$, rather than by independently editing each claim, so that relations among claims and their role in the reasoning chain can affect projection. The projected claims $\widetilde{\mathcal{C}}$ are then synthesized back the rationale level

$$z^{\mathrm{proj}} = \mathcal{S}(\widetilde{\mathcal{C}}), \tag{8}$$

where $\mathcal{S}$ organizes the projected claims into a coherent chain-of-thought. Figure 3 illustrates all four projection types on one real training example, and the projection prompts are given in Appendix **??**.

Screen. The screening stage validates the projected rationale against the deployment observation and ground-truth answer. We define $v = \mathcal{V}(z^{\mathrm{proj}}; x^d, q, \bar{y}) \in \{0, 1\}$, where $v = 1$ only if the projected rationale contains no claims that contradict the deployment observation and the reasoning jointly supports $\bar{y}$; otherwise, $v = 0$ and the candidate is rejected. Screening is performed on the complete rationale rather than on claims independently, since relations among multiple claims may be required to support the final answer.

Loss Function: The student p_θ is fine-tuned using teacher-forced cross-entropy on the projected supervision sequence $w^{\mathrm{proj}} = (z^{\mathrm{proj}}, \bar{y})$, with prompt tokens masked from the loss:

$$\mathcal{L}_{\mathrm{CRP}}(\theta) = -\mathbb{E}_{(x^d, q, z^{\mathrm{proj}}, \bar{y}) \sim \mathcal{D}} \sum_{t=1}^{|w^{\mathrm{proj}}|} \log p_\theta \left(w_t^{\mathrm{proj}} \mid x^d, q, w_{<t}^{\mathrm{proj}} \right). \tag{9}$$

At inference time, the student receives only (x^d, q). The SFT baselines optimize the same objective, replacing z^{proj} with the corresponding supervision target: $\varnothing$ for Answer-only, z^p for Privileged-CoT, and z^d for Deployment-CoT.

5 Analysis: When Does Projection Help?

CRP is motivated by a tradeoff between two forms of information loss. A rationale generated directly from the deployment sensor can omit answer-relevant reasoning in the privileged rationale, while projection can itself lose useful reasoning. We formalize these as the sensor-gap loss and projection loss, and derive the (EPV) that can be computed before training the deployment student to indicate where projection is expected to provide greater benefit.

Sensor gap and projection advantage. Using the privileged rationale z^p as the reference, we define δ as the fraction of answer-relevant reasoning in z^p that is lost in the deployment rationale z^d, and ε as the corresponding fraction lost in the projected rationale z^{proj}. Thus, δ measures the loss across the sensor gap, while ε measures the loss that remains after projection.

Proposition 1 (Projection advantage)**.** *Under the simplified abstract model in Appendix A and the matched supervision conditions specified therein, the difference in expected error between Projected-CoT and Deployment-CoT satisfies*

$$R(\varepsilon) - R(\delta) = c\left(\varepsilon^2 - \delta^2\right), \qquad c > 0, \tag{10}$$

where δ is the rationale information loss across the sensor gap and ε is the rationale information loss remaining after projection. Consequently, Projected-CoT is favored over Deployment-CoT if and only if

$$\varepsilon < \delta.$$

Proof. See Appendix A. □

The proposition gives a direct interpretation of the sensor gap. Projection is useful when it preserves more answer-relevant privileged reasoning than can be retained by a rationale generated from the deployment sensor alone. The reasoning recovered through projection needs to exceed the reasoning lost during projection to bridge the sensor gap.

Table 1: Answer accuracy (%) on RadarQA-300 and AudioQA-300. Values are mean ± standard deviation over seeds. Best results are in **bold** and second-best results are underlined.

	RadarQA-300		AudioQA-300	
Supervision target	**Qwen3.5-4B**	**Qwen3-VL-8B**	**Qwen2.5-Omni-7B**	**Inkling-Small**
Base model	12.3	48.3	25.3	37.5
Answer-only	67.9 ± 1.9	49.6 ± 3.4	56.0 ± 3.9	55.6 ± 3.5
Privileged-CoT	58.9 ± 3.1	52.4 ± 3.1	40.9 ± 2.4	55.6 ± 2.0
Deployment-CoT	70.7 ± 3.2	68.0 ± 4.2	62.3 ± 3.3	56.7 ± 1.9
Projected-CoT (CRP)	**73.8 ± 1.8**	**72.4 ± 1.0**	**68.9 ± 1.8**	**57.5 ± 0.5**

Estimating sensor gap before training. A natural question is whether the projection advantage can be estimated before training a student model for the deployment sensor model. For each paired privileged–deployment sensor sample i, we compute sample-level estimates $\widehat{\delta}_i$ and $\widehat{\varepsilon}_i$ from the corresponding rationales and define the sample-level estimated projection value (EPV) as

$$\mathrm{EPV}_i = \widehat{\delta}_i^2 - \widehat{\varepsilon}_i^2. \tag{11}$$

A positive EPV_i predicts an advantage for Projected-CoT over Deployment-CoT on sample i, while a larger positive value predicts a larger advantage. The analysis suggests two distinct patterns. First, a larger sensor-gap loss, reflected by a larger $\widehat{\delta}_i^2$, creates more opportunity for projection to improve over Deployment-CoT. Second, EPV_i accounts for both this opportunity and the loss that remains after projection, and therefore provides a sample-level estimate of the resulting advantage.

6 Experiments

6.1 Experimental Setup

RadarQA-300: RGB → radar. We construct RadarQA-300 from M4Human (Fan et al., 2026), using synchronized RGB as the privileged observation and millimeter-wave radar in terms of either point cloud or radar heatmap as the deployment sensor observation. The experiment trains Qwen3.5-4B (Qwen Team, 2026) on the released 1,308-instance training set and evaluates on the frozen 300-question RadarQA-300 test set. We additionally evaluate Qwen3-VL-8B (Bai et al., 2025) and InternVL3.5-8B (Wang et al., 2025) for ablation with configurations in Appendix B.

AudioQA-300: audio-visual → audio. We construct AudioQA-300 from MUSIC-AVQA (Li et al., 2022; Liu et al., 2024), using synchronized video and audio as the privileged observation and audio alone as the deployment sensor. We use video-disjoint splits and train the Qwen2.5-Omni-7B model (Xu et al., 2025b) on 1,219 shared training instances before evaluating on the frozen 300-question AudioQA-300 test set. We additionally evaluate Inkling-Small (Thinking Machines Lab, 2026) and full-parameter fine-tuning for ablation study.

Baselines We compare three baselines: Answer-only, Privileged-CoT, and Deployment-CoT, against the proposed Projected-CoT. Within each comparison, the student receives the same deployment observation x^d and question q and uses the same architecture, ground-truth answers, and optimization configuration.

Training and evaluation. Unless otherwise specified, all supervision approaches within a comparison use the same student model, training data, optimization configuration, and checkpoint selection rule. We report answer accuracy, defined as the percentage of questions answered correctly with respect to the ground-truth answer, as mean ± standard deviation over multiple seeds.

7 Results

7.1 Main Results

Table 1 summarizes the results on RadarQA-300 and AudioQA-300.

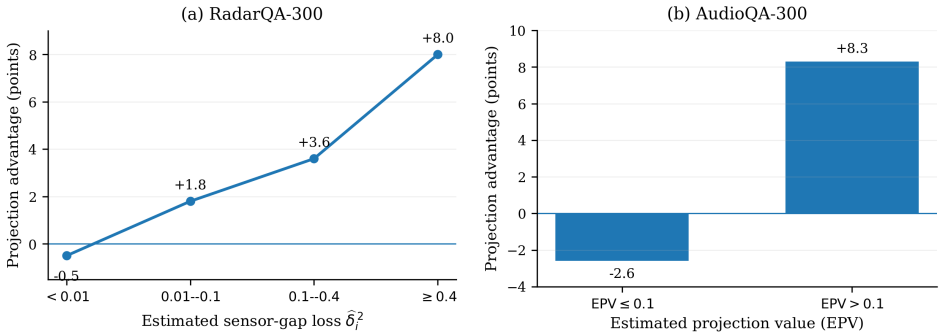

Figure 4: **Empirical analysis of the sensor gap.** Empirical analysis of the sensor gap. (a) On RadarQA-300, the projection advantage increases with the estimated sensor-gap loss. (b) On AudioQA-300, the projection advantage is negative for samples with EPV ≤ 0.1 and positive for samples with EPV > 0.1, consistent with the EPV-guided prediction.

RadarQA-300. Projected-CoT achieves the highest accuracy across both evaluated models. With Qwen3.5-4B, it reaches 73.8%, compared with 67.9% for Answer-only and 70.7% for Deployment-CoT. Directly transferring the privileged rationale can underperform answer-only training, indicating that richer reasoning is not necessarily useful when its claims are not supported by the deployment sensor. Generating rationales directly from radar improves over answer-only supervision, while Projected-CoT provides a further gain, suggesting that privileged reasoning remains useful once it is re-derived against the radar observation. The same pattern holds with Qwen3-VL-8B, where Projected-CoT improves over Deployment-CoT by 4.4 points under full-parameter fine-tuning. A separate Qwen3-VL-8B comparison with GRPO, which uses only final-answer reward without intermediate rationale supervision and optimizes the policy directly, is reported in Appendix E.

AudioQA-300. Projected-CoT also provides consistent gains on the audio-visual-to-audio task. With Qwen2.5-Omni-7B, Answer-only achieves 56.0% accuracy, while Deployment-CoT reaches 62.3%. Projected-CoT further improves accuracy to 68.9%, exceeding Deployment-CoT by 6.6 points and Answer-only by 12.9 points. Privileged-CoT performs substantially worse at 40.9%, again showing that privileged reasoning is not necessarily useful when transferred directly across the sensor gap. The improvement also remains positive with Inkling-Small: Projected-CoT achieves 57.5% compared with 56.7% for Deployment-CoT. Although the margin is smaller in this setting, the ordering between the two rationale-supervision approaches is preserved.

7.2 Empirical Analysis of Sensor Gap

We evaluate the sensor gap analysis of Section 5 in terms of projection advantage and EPV and their indication of performance gain using the radar and audio tasks.

Projection advantage increases with sensor-gap loss. On RadarQA-300, Figure 4 (a) shows that the Projected-CoT advantage over Deployment-CoT increases monotonically across the four selected sample groups characterized by the estimated sensor-gap loss $\widehat{\delta}_i^2$. The margin changes from -0.5 in the lowest-gap group to $+1.8$, $+3.6$, and $+8.0$ as the estimated sensor-gap loss increases. Thus, Projected-CoT provides little benefit when the deployment rationale already retains most of the answer-relevant reasoning, but its advantage becomes substantially larger on samples with greater estimated loss across the sensor gap. This trend is consistent with the predicted dependence of projection advantage on sensor-gap loss.

EPV identifies where projection helps. On the audio evaluation, we partition samples using the predefined threshold EPV $= 0.1$ in Figure 4 (b). For samples with EPV ≤ 0.1, Projected-CoT trails Deployment-CoT by 2.6 percentage points of final-answer accuracy. In contrast, for samples with EPV > 0.1, Projected-CoT leads by 8.3 points, producing a 10.9-point difference between the two partitions. The same ordering is observed across all three matched seeds. This result is consistent with EPV distinguishing samples where projection is expected to recover more useful answer-relevant reasoning than is lost during the projection process.

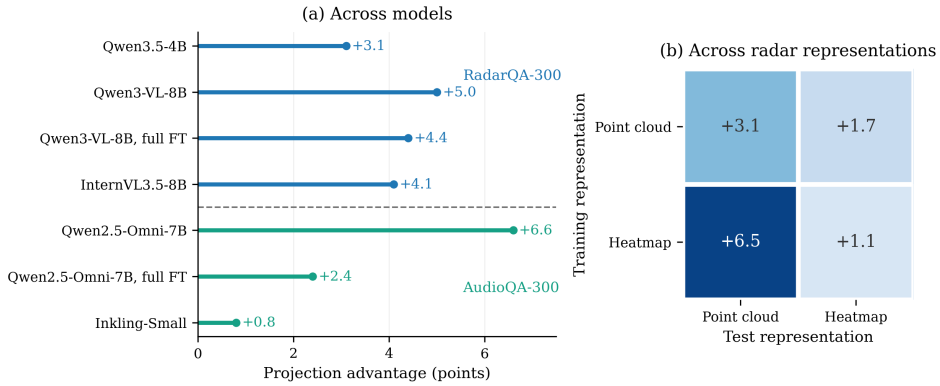

Figure 5: **Projected-CoT remains advantageous across models, fine-tuning regimes, and deployment sensor representations.** (a) Projection advantage across five student backbones, two deployment modalities, and LoRA and full-parameter fine-tuning. (b) Projection advantage on RadarQA-300 when training and test using the same or different radar representations, either point clouds or radar heatmaps.

7.3 Ablations across Models, Fine Training, and Sensor Representations

We further examine whether the Projected-CoT advantage over Deployment-CoT persists across different student models, fine-tuning regimes, and deployment sensor representations.

Models and fine-tuning regimes. Figure 5(a) shows that the advantage extends across additional model backbones and training settings. On RadarQA300, Projected-CoT improves final-answer accuracy over Deployment-CoT by 4.1 points with InternVL3.5-8B and by 5.0 points with Qwen3-VL-8B under low-rank fine-tuning, i.e., LoRA. The advantage remains 4.4 points when Qwen3-VL-8B is trained with full-parameter fine-tuning. On AudioQA300, full-parameter fine-tuning of Qwen2.5-Omni-7B yields a 2.4-point gain, while Inkling-Small yields a 0.8-point gain. These results show that the improvement is not specific to a particular student backbone or fine-tuning regime.

Radar sensor representations. Figure 5(b) evaluates the effect in the radar representation between training and testing. Projected-CoT improves final-answer accuracy over Deployment-CoT by 1.1 points when both training and testing use radar heatmaps, by 1.7 points when training uses point clouds and testing uses radar heatmaps, and by 6.5 points when training uses radar heatmaps and testing uses point clouds. The positive margins under both matched and mismatched representations indicate that the projection advantage is not tied to a single radar input format. The detailed records of final answer accuracy for both methods are included in Appendix C.

8 Limitations

Radar ground-truth answers are derived from pose and trajectory annotations, while audio answers use automated consensus. The two benchmarks, RadarQA-300 and AudioQA-300, are limited to specific data collection settings, including empty rooms and limited background noise. Our sensor-gap analysis relies on a simplified abstract model that linearizes the contribution of rationale components to the final answer. Finally, the deployment sensors are limited to radar and audio.

9 Conclusion

We introduced Cross-sensor Rationale Projection (CRP) to transfer privileged rationale supervision across sensor gaps by re-deriving rationale claims against the deployment observation before distillation. On RadarQA-300 and AudioQA-300, Projected-CoT consistently outperforms final-answer-only, Privileged-CoT, and Deployment-CoT supervision, with gains of 5.9 and 12.9 points over final-answer-only training, respectively. Our theoretical analysis further motivates the EPV, which consistently predicts where the empirical advantage of projected supervision is larger.

AI USE STATEMENT

Generative AI tools were used to assist with method implementation, dataset processing, qualitative analysis, and interpretation of results. They were also used for figure plotting, code generation and editing, literature search and summarization, manuscript editing, and reference preparation. The authors reviewed all AI-assisted work and were substantially involved in verifying and finalizing the methods, analyses, text, claims, code, figures, and other artifacts produced with the aid of generative AI. The authors take responsibility for the final content of this work.

ETHICS STATEMENT

Our benchmarks are constructed from the public M4Human and MUSIC-AVQA datasets and follow their respective data-use and privacy provisions. M4Human reports voluntary participant recruitment, facial de-identification, and that participants were informed of the public release of de-identified data for research purposes. MUSIC-AVQA uses publicly available YouTube videos and reports that the dataset contains no personally identifiable information.

REPRODUCIBILITY STATEMENT

We will publicly release RadarQA-300 and AudioQA-300, together with the CRP software and model checkpoints used to reproduce the reported results on both benchmarks. The release will also include the benchmark construction and evaluation scripts needed to reproduce the experimental protocol.

REFERENCES

Sizhe An, Yin Li, and Umit Ogras. mri: Multi-modal 3d human pose estimation dataset using mmwave, rgb-d, and inertial sensors. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 27414–27426. Curran Associates, Inc., 2022. doi: 10.52202/068431-1988. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/af9c9c6d2da701da5a0acf91ec217815-Paper-Datasets_and_Benchmarks.pdf.

Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, Wenbin Ge, Zhifang Guo, Qidong Huang, Jie Huang, Fei Huang, Binyuan Hui, Shutong Jiang, Zhaohai Li, Mingsheng Li, Mei Li, Kaixin Li, Zicheng Lin, Junyang Lin, Xuejing Liu, Jiawei Liu, Chenglong Liu, Yang Liu, Dayiheng Liu, Shixuan Liu, Dunjie Lu, Ruilin Luo, Chenxu Lv, Rui Men, Lingchen Meng, Xuancheng Ren, Xingzhang Ren, Sibo Song, Yuchong Sun, Jun Tang, Jianhong Tu, Jianqiang Wan, Peng Wang, Pengfei Wang, Qiuyue Wang, Yuxuan Wang, Tianbao Xie, Yiheng Xu, Haiyang Xu, Jin Xu, Zhibo Yang, Mingkun Yang, Jianxin Yang, An Yang, Bowen Yu, Fei Zhang, Hang Zhang, Xi Zhang, Bo Zheng, Humen Zhong, Jingren Zhou, Fan Zhou, Jing Zhou, Yuanzhi Zhu, and Ke Zhu. Qwen3-vl technical report, 2025. URL https://arxiv.org/abs/2511.21631.

Hardy Chen, Haoqin Tu, Fali Wang, Hui Liu, Xianfeng Tang, Xinya Du, Yuyin Zhou, and Cihang Xie. Sft or rl? an early investigation into training r1-like reasoning large vision-language models. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=wZI5qkQeDF.

Yuhao Dong, Zuyan Liu, Hai-Long Sun, Jingkang Yang, Winston Hu, Yongming Rao, and Ziwei Liu. Insight-v: Exploring long-chain visual reasoning with multimodal large language models, 2025. URL https://arxiv.org/abs/2411.14432.

Junqiao Fan, Yunjiao Zhou, Yizhuo Yang, Xinyuan Cui, Jiarui Zhang, Lihua Xie, Jianfei Yang, Chris Xiaoxuan Lu, and Fangqiang Ding. M4human: A large-scale multimodal mmwave radar benchmark for human mesh reconstruction, 2026. URL https://arxiv.org/abs/2512.12378.

Mingda Han, Huanqi Yang, Zehua Sun, Wenhao Li, Yanni Yang, Guoming Zhang, Yetong Cao, Weitao Xu, and Pengfei Hu. Ragent: Physics-aware agentic reasoning for training-free mmwave human activity recognition, 2026. URL https://arxiv.org/abs/2603.27571.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Fushuo Huo, Wenchao Xu, Jingcai Guo, Haozhao Wang, and Song Guo. C2kd: Bridging the modality gap for cross-modal knowledge distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16006–16015, June 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Huo_C2KD_Bridging_the_Modality_Gap_for_Cross-Modal_Knowledge_Distillation_CVPR_2024_paper.html.

Zengyuan Lai, Jiarui Yang, Songpengcheng Xia, Lizhou Lin, Lan Sun, Renwen Wang, Jianran Liu, Qi Wu, and Ling Pei. Radarllm: Empowering large language models to understand human motion from millimeter-wave point cloud sequence. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2026. URL https://arxiv.org/abs/2504.09862.

Maksim Lapin, Matthias Hein, and Bernt Schiele. Learning using privileged information: SVM+ and weighted SVM. *Neural Networks*, 53:95–108, 2014. doi: 10.1016/j.neunet.2014.02.002.

Guangyao Li, Yake Wei, Yapeng Tian, Chenliang Xu, Ji-Rong Wen, and Di Hu. Learning to answer questions in dynamic audio-visual scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19108–19118, 2022. URL https://arxiv.org/abs/2203.14072.

Xiulong Liu, Zhikang Dong, and Peng Zhang. Tackling data bias in music-avqa: Crafting a balanced dataset for unbiased question-answering. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 4478–4487, 2024. URL https://arxiv.org/abs/2310.06238.

David Lopez-Paz, Leon Bottou, Bernhard Schölkopf, and Vladimir Vapnik. Unifying distillation and privileged information. In *International Conference on Learning Representations*, 2016. URL https://arxiv.org/abs/1511.03643.

Qwen Team. Qwen3.5-4B model card, March 2026. URL https://huggingface.co/Qwen/Qwen3.5-4B. Official model card.

Hao Shao, Shengju Qian, Han Xiao, Guanglu Song, Zhuofan Zong, Letian Wang, Yu Liu, and Hongsheng Li. Visual cot: Advancing multi-modal language models with a comprehensive dataset and benchmark for chain-of-thought reasoning, 2024a. URL https://arxiv.org/abs/2403.16999.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models, 2024b. URL https://arxiv.org/abs/2402.03300.

Thinking Machines Lab. Inkling-Small model card, July 2026. URL https://thinkingmachines.ai/model-card/inkling-small/. Official model card; open-weight release.

Kanghui Tian, Siyuan Liu, Ziang Yan, Sheng Xia, Shuai Dong, and Yi Wang. Vicur: Visual cues as recoverable privilege for multimodal on-policy distillation, 2026. URL https://arxiv.org/abs/2606.05718.

Vladimir Vapnik and Akshay Vashist. A new learning paradigm: Learning using privileged information. *Neural Networks*, 22(5–6):544–557, 2009. doi: 10.1016/j.neunet.2009.06.042.

Weiyun Wang, Zhangwei Gao, Lixin Gu, Hengjun Pu, Long Cui, Xingguang Wei, Zhaoyang Liu, Linglin Jing, Shenglong Ye, Jie Shao, et al. Internvl3.5: Advancing open-source multimodal models in versatility, reasoning, and efficiency. *arXiv preprint arXiv:2508.18265*, 2025. URL https://arxiv.org/abs/2508.18265.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2023. URL https://arxiv.org/abs/2201.11903.

Xueqing Wu, Yuheng Ding, Bingxuan Li, Pan Lu, Da Yin, Kai-Wei Chang, and Nanyun Peng. Visco: Benchmarking fine-grained critique and correction towards self-improvement in visual reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9527–9537, June 2025.

Guowei Xu, Peng Jin, Ziang Wu, Hao Li, Yibing Song, Lichao Sun, and Li Yuan. Llava-cot: Let vision language models reason step-by-step, 2025a. URL https://arxiv.org/abs/2411.10440.

Jin Xu, Zhifang Guo, Jinzheng He, Hangrui Hu, Ting He, Shuai Bai, Keqin Chen, Jialin Wang, Yang Fan, Kai Dang, Bin Zhang, Xiong Wang, Yunfei Chu, and Junyang Lin. Qwen2.5-omni technical report. *arXiv preprint arXiv:2503.20215*, 2025b. URL https://arxiv.org/abs/2503.20215.

Kangning Zhang, Yixing Li, Shuai Shao, Qingyao Li, Zhengxi Lu, Zhiyuan Yao, Jianghao Lin, Wenxiang Jiao, Yuan Lu, Weiwen Liu, Weinan Zhang, and Yong Yu. Vad: Attributing visual evidence for target reconstruction in multimodal on-policy distillation, 2026. URL https://arxiv.org/abs/2607.28590.

Zihan Zhao, Kaushik Pendiyala, Masood Mortazavi, and Ning Yan. Pulse: Privileged knowledge transfer from rich to deployable sensors for embodied multi-sensory learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 3946–3954, June 2026. URL https://openaccess.thecvf.com/content/CVPR2026W/SenseOfSpace/html/Zhao_PULSE_Privileged_Knowledge_Transfer_from_Rich_to_Deployable_Sensors_for_CVPRW_2026_paper.html.