PACT: TOWARD VERIFIABLE PROACTIVE ASSISTANCE FOR WEARABLE LANGUAGE AGENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

A wearable agent that continuously observes its user must decide, without an explicit request, whether to intervene, what assistance to provide, and how to deliver it under device and timing constraints. Existing training largely relies on answer-level supervision, which provides target decisions but does not retain an executable mechanism for verifying new policy outputs. Group-relative reinforcement learning poses a second difficulty, as identical rollouts receive identical rewards and yield no advantage signal. We formalize *label derivability* as the property that the decision mechanism remains executable after data construction, and propose PACT, a training framework that exploits this property for proactive policy learning. PACT derives supervision from executable decision rules over measurable conditions, re-executes these rules to verify policy outputs, and forms optimization groups across label-changing counterfactual pairs to preserve reward contrast when within-input variation disappears. We further introduce POISE, a benchmark of 8,559 wearable decisions across twelve scenarios, 49 tools, and nine delivery channels, grounded in sensor and contextual observations. Evaluation on POISE reveals persistent difficulty in distinguishing warranted intervention from abstention among prompting and inference-time methods. PACT improves over supervised fine-tuning and alternative RL objectives on POISE, while also yielding gains in tool selection and proactivity-score calibration on external benchmark.

1 INTRODUCTION

A cyclist's wearable detects a vehicle rapidly approaching from behind (Fig. 1, left). A proactive agent must decide *whether* to intervene, *what* assistance to provide, and *how* to deliver it under device and timing constraints. Wearable devices are becoming a natural platform for language agents that continuously observe users and may initiate assistance without explicit requests (Kim et al., 2024). Recent proactive-agent benchmarks commonly rely on answer-level supervision, where a model directly judges whether an intervention is appropriate (Yang et al., 2026; Lu et al., 2025; Liu et al., 2025). Such supervision records the final decision but not the variables that justify it. A correct prediction may therefore reflect either decisive evidence or correlated context, which the final label cannot distinguish. Because many wearable situations require no intervention, aggregate accuracy also provides limited evidence that a policy responds selectively to conditions that warrant action. These limitations arise from the supervision rather than from any particular model.

Post-hoc diagnostics can expose such failures through controlled perturbations (Gardner et al., 2020) or additional annotations (Ribeiro et al., 2020). However, diagnostics added after data construction do not necessarily preserve the mechanism that produced the original label. An annotation may identify relevant factors such as approach speed, physiological state, channel availability, or latency without specifying how they jointly determine the intervention. Likewise, an authored counterfactual must be adjudicated again to establish whether its label changes. If this judgment is delegated to a learned reward model during training, policy optimization may exploit the model rather than satisfy the underlying conditions (Gao et al., 2023) (Fig. 1a). An executable decision rule is therefore required to derive labels from measurable conditions and enable subsequent verification.

Wearable proactive assistance is particularly amenable to executable supervision because many interventions depend on measurable conditions such as approach speed, physiological deviation, channel availability, and timing constraints. We refer to this property as *label derivability*, where the

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.

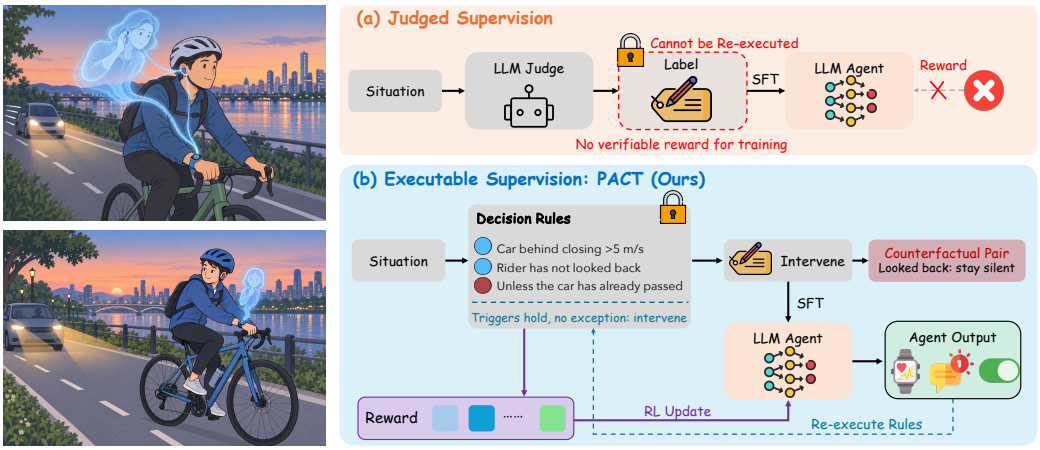

Figure 1: Judged versus executable supervision for proactive wearable agents. Left: a single decisive variable flips the correct intervention. (a) Judged labels. (b) Labels derived from executable decision rules, which PACT reuses to verify policy outputs.

mechanism determining an intervention is specified before data construction, retained throughout the process, and re-executed against new policy outputs. The same mechanism can identify the decisive variable, generate counterfactuals with guaranteed label changes, and verify whether, what, and how a policy intervenes. This resolves the supervision side, but group-relative optimization introduces a separate difficulty. When a policy produces identical behavior across rollouts of the same prompt, group-relative objectives (Shao et al., 2024) receive identical rewards and provide no advantage signal (Yu et al., 2025). Effective learning therefore also requires an optimization strategy that preserves reward contrast when within-prompt variation disappears.

To address these challenges, we propose **PACT**, a training framework for learning proactive policies from executable decision rules (Fig. 1b). PACT uses measurable conditions to supervise whether an intervention is warranted, what assistance is appropriate, and how it should be delivered. Because labels are derived from these conditions rather than judged from generated text, decisive variables can be varied to form counterfactual pairs with guaranteed label changes. PACT re-executes the same rules on policy outputs to provide verifiable rewards and couples paired situations during group-relative optimization to preserve reward contrast when within-prompt behavior becomes identical. The same rules therefore support label supervision, counterfactual construction, and policy optimization. We further introduce **POISE**, a benchmark of 8,559 wearable decisions spanning diverse sensing contexts, intervention actions, delivery channels, and timing constraints. Our contributions:

- We formalize *label derivability* as the property that the mechanism determining an intervention remains executable after data construction, allowing the same rules to support label assignment, counterfactual construction, and verifiable policy evaluation.
- We propose **PACT**, a training framework that re-executes decision rules to verify policy outputs and couples counterfactual pairs during group-relative optimization to preserve reward contrast when within-prompt variation vanishes.
- We introduce **POISE**, a benchmark of 8,559 wearable decisions covering twelve scenarios, 49 tools, and nine delivery channels, grounded in sensor and contextual observations to evaluate whether, what, and how an agent should intervene.
- Empirical results characterize **POISE** as a challenging benchmark and show that PACT improves proactive decision making over supervised and alternative RL rewards, while also yielding gains in tool selection and proactivity-score calibration on an external benchmark.

2 Related Work

Context-Aware Language Agents. Language agents are increasingly deployed in settings where inputs include continuous sensor and contextual signals rather than explicit user requests (Kim et al., 2024; Xu et al., 2024), building on prior work in context-aware (Abowd et al., 1999; Dey, 2001)

and mixed-initiative computing (Horvitz, 1999). Recent methods use such signals to determine whether intervention is warranted and which tools or actions to invoke (Deng et al., 2023; Yang et al., 2026; Lu et al., 2025; Li et al., 2025), alongside advances in tool-using agents (Schick et al., 2023; Yao et al., 2022). Some approaches instead generate candidate utterances internally and decide whether to surface them to the user (Liu et al., 2025). These methods mainly supervise intervention, content, and tool use, while delivery choices such as device, modality, intensity, and timing typically remain outside the prediction target despite their effect on interruption cost (Mehrotra et al., 2016). Labels are also typically assigned at the response level rather than derived from a retained and executable decision mechanism. Contrast sets (Gardner et al., 2020) and counterfactually augmented data (Kaushik et al., 2019) can probe sensitivity to contextual changes, but their perturbations are generally constructed independently of the original executable labeling rule.

Agentic Reinforcement Learning. Reinforcement learning is widely used to optimize language models and agents after supervised fine-tuning (Ouyang et al., 2022; Guo et al., 2025). Directly checkable correctness, as in mathematical verification or executable tests, lets a verifier provide the reward without a learned reward model (Shao et al., 2024; Guo et al., 2025). Group-relative objectives optimize policies by normalizing rewards across rollouts from the same prompt (Shao et al., 2024). In more general agentic tasks, however, supervision is often sparse or available only at the outcome level, motivating learned reward models (Ouyang et al., 2022) and process reward models that score intermediate reasoning steps (Lightman et al., 2024). Because these rewards are learned, they may introduce approximation error and can be exploited by the policy (Gao et al., 2023). Group-relative optimization also becomes uninformative when all rollouts within a group receive the same reward, since the normalized advantage vanishes (Yu et al., 2025). PACT addresses these two limitations with executable decision rules that directly verify policy outputs and counterfactual pairing that preserves reward contrast when within-prompt variation disappears.

3 METHODOLOGY

Problem Formulation. A proactive situation is represented by a structured state $\mathbf{s} \in \mathcal{S}$ containing sensor and contextual variables. The agent observes $x = (g(\mathbf{s}), p, a)$, where $g(\mathbf{s})$ renders the observable state as text, p denotes the user persona, and a the available delivery channels. It produces a structured output $o = (e, \sigma, T, \Pi)$ containing decision evidence, a proactivity score, tool calls, and a delivery plan (Fig. 2, top). Each scenario defines an executable specification $\mathcal{C}$ whose trigger and suppression predicates determine the intervention label $y(\mathbf{s})$ and whose constraints specify valid content, delivery, and timing. Executing the same specification on a policy output induces a verifiable reward $R(o, \mathbf{s})$ over whether, what, and how the policy intervenes. We learn $\pi_\theta(o \mid x)$ by

$$\max_\theta \ \mathbb{E}_{\mathbf{s}\sim\mathcal{D},\, o\sim\pi_\theta(\cdot|x)} \left[R(o, \mathbf{s})\right], \tag{1}$$

where R is deterministically evaluated from $\mathcal{C}$ and requires no learned reward model.

Overview. Figure 2 illustrates the overall workflow of PACT. Executable decision rules first derive structured targets and counterfactual pairs from sampled states, which are used to initialize the proactive policy with supervised learning. PACT then re-executes the same rules on policy rollouts to obtain verifiable rewards and computes group-relative advantages over pair-contrastive groups for reinforcement learning. The executable rules therefore provide a common basis for supervision, verification, and policy optimization.

Executable Supervision. Each scenario is associated with an executable specification $\mathcal{C}$ that defines a distribution over structured states, trigger and suppression predicates, and valid action and delivery constraints. For a sampled state $\mathbf{s}$, the intervention label $y(\mathbf{s})$ and structured target $o^\star(\mathbf{s})$ are derived directly from $\mathcal{C}$, while the observation $g(\mathbf{s})$ is rendered from the state without conditioning on the label. We construct a counterfactual pair $(\mathbf{s}^+, \mathbf{s}^-)$ by varying decisive variables while holding the remaining state fixed such that $y(\mathbf{s}^+) \neq y(\mathbf{s}^-)$. The set of changed variables, $V^\star$, provides a controlled contrast for evaluation and optimization. The derived targets initialize the policy through supervised fine-tuning (SFT) using

$$\mathcal{L}_{\mathrm{SFT}} = -\mathbb{E}_{(\mathbf{s},x)} \left[\log \pi_\theta(o^\star(\mathbf{s}) \mid x)\right], \tag{2}$$

yielding π_{ref} as both the initialization and KL reference for subsequent reinforcement learning.

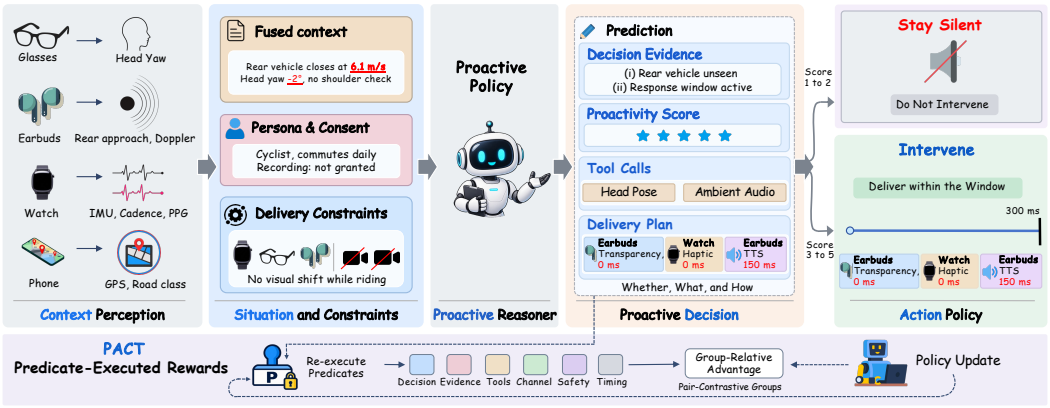

Figure 2: Overview of PACT. Top: the policy maps wearable context, persona and delivery constraints to a structured decision. Bottom: the construction predicates are re-executed on each rollout to produce six reward components, normalized over pair-contrastive groups.

Predicate-Executed Reward. For each rollout o under state $\mathbf{s}$, PACT re-executes the executable specification to evaluate six verifiable reward components (Fig. 2, bottom) and aggregates the applicable components as

$$R(o, \mathbf{s}) = \frac{\sum_{k \in \mathcal{K}(o,\mathbf{s})} w_k r_k(o, \mathbf{s})}{\sum_{k \in \mathcal{K}(o,\mathbf{s})} w_k}, \tag{3}$$

where each $r_k \in [0, 1]$ is computed deterministically from the policy output and specification, and $\mathcal{K}(o, \mathbf{s})$ indexes the components applicable to the sample. The decision component scores the ordinal proactivity prediction and its induced intervention decision. The evidence component checks whether the cited evidence identifies the decisive variable, the tool component evaluates the selected tool set, and the channel component evaluates the delivery device and modality. The safety component enforces consent and channel constraints, while the timing component evaluates compliance with the actionable latency window. Exact definitions and weights are provided in Appendix B.3.

Pair-Contrastive Grouping. Group-relative objectives estimate advantages by normalizing rewards across rollouts of the same input. When these rollouts collapse to identical behavior, they receive identical rewards and the within-group advantage vanishes entirely. PACT instead forms each optimization group from rollouts of a controlled counterfactual pair, $G = G(x^+) \cup G(x^-)$. Because the paired situations differ in a decisive variable and have different derived labels, identical behavior across them can still induce distinct rewards. Pair-contrastive grouping therefore preserves an informative reward contrast even when behavior within an input collapses.

Optimization. Given the pair-contrastive groups above, we optimize the standard clipped group-relative objective with a KL penalty to the supervised policy π_{ref} (Schulman et al., 2017; Shao et al., 2024). For rollout o_i with group-normalized advantage A_i, let $\rho_{i,t}(\theta) = \pi_\theta(o_{i,t} \mid x, o_{i,<t}) / \pi_{\theta_{\text{old}}}(o_{i,t} \mid x, o_{i,<t})$. The objective can be formulated as:

$$\mathcal{J}(\theta) = \mathbb{E}\left[\frac{1}{|G|}\sum_{i \in G}\frac{1}{|o_i|}\sum_{t=1}^{|o_i|}\min(\rho_{i,t}A_i,\ \text{clip}(\rho_{i,t}, 1-\epsilon, 1+\epsilon)A_i)\right] - \beta D_{\text{KL}}(\pi_\theta \,\|\, \pi_{\text{ref}}), \tag{4}$$

where A_i is obtained by normalizing the predicate-executed rewards in Eq. 3 within the pair-contrastive group G. PACT leaves the underlying group-relative objective unchanged, modifying only the reward source and the composition of the normalization group.

4 The POISE Benchmark

POISE evaluates proactive wearable agents across the complete intervention decision: whether to act, what assistance to provide, and how to deliver it under device and timing constraints. It contains 8,559 decisions across twelve everyday scenarios, with each label retained as an executable

specification that supports counterfactual evaluation and policy verification. Fig. 3 summarizes its coverage, while Table 1 compares POISE with the closest proactive-agent benchmarks.

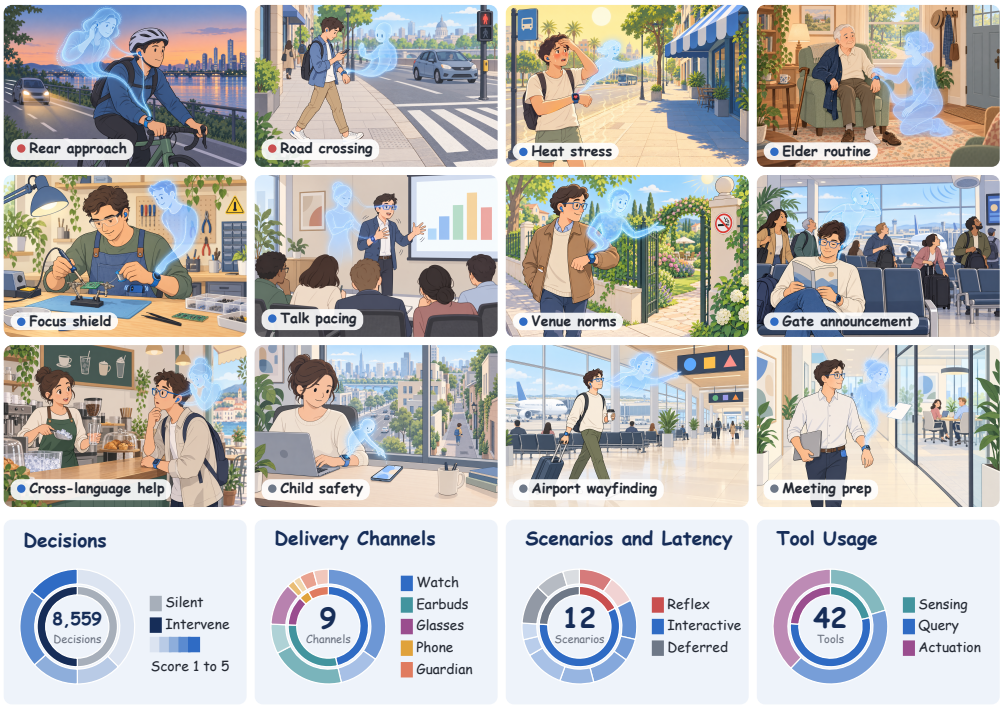

Figure 3: POISE overview. Twelve proactive wearable scenarios and their distribution across decisions, delivery channels, latency, and tools. Demo samples are available at this link.

Task and Evaluation Space. Each example presents a sensory context, user persona, and available delivery channels, and requires a structured decision along three dimensions. The agent determines *whether* to intervene through a graded proactivity score, *what* assistance to provide through calls to a catalog of 49 sensing, query, and actuation tools, 42 of which appear in reference decisions, and *how* to deliver it through a device, modality, intensity, and timing plan. POISE additionally evaluates decision evidence, safety constraints, and valid abstention or escalation. Unlike existing benchmarks that primarily supervise intervention decisions and content, POISE explicitly evaluates the full delivery process under available channel and timing constraints.

Table 1: Comparison with existing proactive-agent benchmarks. Tools: distinct tools in reference decisions. Reusable: counterfactual pairs or executable rewards. RM: reward model.

Benchmark	Modalities	Samples	Tools	Supervision		Label source	Reusable	
				Content	Delivery		Pairs	Reward
ProactiveBench	Desktop events	7,023	✗	✓	✗	Human + RM	✗	✗
CAB-Lite	Vision, audio	300	16	✓	✗	LLM	✗	✗
CAB (Full)		1,000	20					
POISE-Lite	IMU, PPG, GPS, audio	494	**42**	✓	✓	Executable decision rules	247	✓
POISE (Full)		**8,559**					**4,282**	

Coverage and Data Grounding. POISE spans twelve scenarios covering mobility, health, dependent care, focused work, travel, and social interaction. Scenarios vary in latency requirements, protected parties, available channels, and escalation constraints, creating distinct intervention policies even when the observed context is similar. Intervention and abstention cases are approximately balanced to prevent trivial always-act or always-silent strategies. Inputs include recorded wearable activity (Reiss & Stricker, 2012), indoor positioning (Torres-Sospedra et al., 2014), photoplethysmography (Reiss et al., 2019), outdoor trajectories (Zheng et al., 2010), and speech (Panayotov et al., 2015), alongside contextual records for announcements, cross-language assistance, and meeting preparation. These sources expose agents to both measured sensor dynamics and contextual constraints that affect proactive decisions.

Executable Diagnostics. Executable specifications allow POISE to evaluate more than answer accuracy. Each counterfactual pair differs in decisive variables and has a different derived label, enabling pair consistency to measure whether a model responds to the relevant evidence rather than correlated context. The same specification verifies tool use, delivery, safety, and timing, providing consistent supervision for both evaluation and reinforcement learning. POISE further includes POISE-Lite, a scenario-balanced subset for efficient model comparison. Dataset splits keep each counterfactual pair within the same split to prevent leakage between training and evaluation. Three expert annotators independently assess the faithfulness of the rendered narratives, with construction details, annotation procedures, and full statistics provided in Appendix A.

5 EXPERIMENTS

Benchmarks and Metrics. We evaluate on the POISE test split and POISE-Lite, and on the CAB test split of 400 samples. The CAB-comparable metrics are decision accuracy (Acc-P), missed and false intervention rates (MD, FD), score RMSE and tool precision, recall and F1. POISE adds pair consistency (Pair-C.), the share of counterfactual pairs whose two halves are both decided correctly. A policy that ignores the input cannot raise it. POISE further reports channel F1 over device and modality, the safety-violation rate (MSVR), decisive-variable identification, actionability within the window, and abstention. The POISE score is the mean of Acc-P, Pair-C., Tool-F1 and Channel-F1. Three floor policies that never read the input are scored under the same protocol. Appendix A gives the split statistics and the formal definition of every metric. Rates and bounded scores are reported $\times 100$ throughout, and RMSE is reported on the original scale.

Baselines. We compare prompting and training baselines. Prompting baselines include open-weight models under the Direct protocol, seven prompt variants of DeepSeek-V3.2 (DeepSeek-AI et al., 2025), and three input-agnostic floor policies. We also evaluate seven training-free methods on top of Direct: chain of thought (CoT) (Wei et al., 2022), self-consistency with five samples (Wang et al., 2023), ReAct (Yao et al., 2022), Reflexion (Shinn et al., 2023), Inner Thoughts (Liu et al., 2025), Proactive Agent (Lu et al., 2025), and native reasoning at high effort. Training baselines use full-parameter supervised fine-tuning on the derived targets or GRPO (Shao et al., 2024), initialized from either the base model or the supervised checkpoint. GRPO uses three reward variants. The outcome reward scores only the proactivity decision, the judge reward additionally scores proactivity and tool selection when the intervention decision is correct, and the PACT reward evaluates all six verifiable components. Base models are evaluated zero-shot using Direct on POISE and the CAB zero-shot protocol on CAB. Appendix B.1 details the introduction of each baseline.

Implementation Details. We train Qwen2.5-1.5B-Instruct (Qwen et al., 2025) and Qwen3-1.7B (Yang et al., 2025) for the main comparisons using the same two-stage procedure and hyperparameters. The backbone-scaling comparison additionally uses Qwen2.5-3B-Instruct and Qwen2.5-7B-Instruct under the same SFT and PACT configurations. SFT updates all parameters with LLaMA-Factory (Zheng et al., 2024) for five epochs at a learning rate of 10^{-5}, with the persona included in the input and decision evidence in the target. RL initializes from the supervised checkpoint and applies token-level GRPO with Eq. 3 using the EasyR1 fork of verl (Sheng et al., 2025). Training runs for two epochs with 16 prompts per step, five rollouts per prompt at temperature 1.0, a learning rate of 10^{-5}, a KL coefficient of $\beta = 0.01$ against the supervised policy, and a clipping range of $\varepsilon = 0.2$. The outcome, judge, and PACT reward variants share this configuration and differ only in their reward definitions. The sensitivity sweep varies β, evidence weight, clipping range, or rollouts per input, one at a time. Remaining settings are in Appendix B.3. Prompted baselines use vLLM (Kwon et al., 2023) under the Direct protocol, with one call per example at temperature zero and native reasoning disabled when configurable. All experiments run on an NVIDIA A100 80GB GPU.

5.1 MAIN RESULTS

Prompted Models. Table 2 shows that prompting rarely produces selective intervention on POISE. The highest open-weight POISE-Lite score is 39.0, below the random floor of 42.8. Most prompted models intervene on nearly every sample, and several lose performance on the full scenario distribution. This decline is consistent with difficulty applying one prompt across heterogeneous suppression conditions. Larger backbones do not consistently help. Low pair consistency

suggests plausible single-case actions without reliable sensitivity to the condition that changes the correct decision, even when the surrounding situation remains similar.

Table 2: Prompted open-weight models on POISE-Lite and POISE. Bold and underlined values mark the best and second-best scores per split.

		POISE-LITE						POISE (FULL)					
Model	Size	Acc-P↑	FD↓	Pair-C.↑	Tool-F1↑	Chan.-F1↑	POISE↑	Acc-P↑	FD↓	Pair-C.↑	Tool-F1↑	Chan.-F1↑	POISE↑
Qwen3-0.6B	0.6B	50.0	100.0	0.0	1.3	18.9	17.6	50.0	99.9	0.1	2.2	20.3	18.1
Qwen3-1.7B	1.7B	50.4	99.2	0.8	6.6	27.3	21.3	50.2	99.7	0.4	7.1	28.4	21.5
Qwen3-4B	4B	50.2	99.2	0.8	13.6	28.9	23.4	50.3	99.4	0.6	13.6	29.2	23.4
Qwen3-4B-Inst.-2507	4B	50.2	99.6	0.4	18.0	30.1	24.7	50.3	99.3	0.7	<u>19.4</u>	31.3	25.4
Qwen3-8B	8B	50.4	98.8	1.2	13.1	27.6	23.1	49.3	95.4	4.0	10.5	<u>33.1</u>	24.3
Qwen2.5-1.5B-Inst.	1.5B	49.0	**29.6**	10.5	**37.6**	<u>42.7</u>	<u>34.9</u>	49.7	98.7	1.2	2.6	24.6	19.5
Qwen2.5-3B-Inst.	3B	<u>54.7</u>	<u>41.7</u>	**24.7**	<u>33.5</u>	**43.2**	**39.0**	<u>50.9</u>	97.3	2.7	3.9	18.9	19.1
Qwen2.5-7B-Inst.	7B	**58.5**	72.1	<u>21.5</u>	18.7	33.8	33.1	**54.7**	**78.0**	**14.8**	17.9	**33.7**	**30.3**
Llama-3.2-3B-Inst.	3B	50.0	100.0	0.0	6.6	24.6	20.3	50.1	99.8	0.2	5.9	26.8	20.8
Llama-3.1-8B-Inst.	8B	50.4	96.8	2.4	14.9	29.2	24.2	50.6	96.1	3.3	14.6	29.4	24.5
Gemma-3-4B-it	4B	50.0	99.6	0.4	11.5	25.8	21.9	50.0	99.4	0.6	11.7	28.0	22.6
GLM-4-9B-0414	9B	49.8	93.5	6.5	12.9	29.5	24.6	50.4	<u>94.6</u>	<u>4.7</u>	**30.5**	31.5	<u>29.3</u>
Always silent	–	50.0	0.0	0.0	50.0	50.0	37.5	50.0	0.0	0.0	50.0	50.0	37.5
Always act	–	50.0	100.0	0.0	50.0	50.0	37.5	50.0	100.0	0.0	50.0	50.0	37.5
Random	–	50.0	52.2	21.1	50.0	50.0	42.8	51.7	47.1	26.2	50.0	50.0	44.5

Inference-Time Methods. Table 3 shows that training-free inference methods offer uneven gains on POISE. On POISE-Lite, Proactive Agent raises Qwen3-4B POISE by 15.2 points, while decision accuracy rises by only 1.8 points. Its benefit does not carry over to Qwen2.5-7B, where Direct remains competitive. Other methods vary by backbone or stay near Direct. This pattern suggests that extra inference steps change actions without reliably setting an intervention boundary.

Table 3: Training-free methods on POISE. Bold marks the best score per model.

		POISE-LITE						POISE (FULL)					
Model	Method	Acc-P↑	FD↓	Pair-C.↑	Tool-F1↑	Chan.-F1↑	POISE↑	Acc-P↑	FD↓	Pair-C.↑	Tool-F1↑	Chan.-F1↑	POISE↑
Qwen3-8B	Direct	50.4	98.8	1.2	13.1	27.6	23.1	49.3	95.4	4.0	10.5	**33.1**	24.3
	CoT	51.8	96.4	3.6	15.1	27.6	24.5	51.4	96.6	3.4	15.1	29.2	24.8
	SC	50.4	98.8	1.2	12.6	28.0	23.1	50.4	99.0	0.9	13.4	29.0	23.4
	ReAct	50.6	98.0	2.0	13.6	27.1	23.3	50.7	98.2	1.6	14.4	24.3	22.8
	Reflexion	50.6	98.4	1.6	13.1	27.9	23.3	50.2	99.3	0.6	14.0	29.3	23.5
	Inner Th.	50.0	100.0	0.0	13.3	27.1	22.6	50.0	100.0	0.0	14.0	28.4	23.1
	Proact.	**54.5**	**87.0**	**11.3**	**23.1**	**31.1**	**30.0**	**53.3**	**88.4**	**9.0**	**22.2**	31.4	**29.0**
	Reason.-high	51.6	95.1	4.5	16.5	24.1	24.2	51.3	96.9	3.0	16.8	25.4	24.1
Qwen3-4B	Direct	50.2	99.2	0.8	13.6	28.9	23.4	50.3	99.4	0.6	13.6	29.2	23.4
	CoT	51.8	92.7	5.7	15.5	31.5	26.1	52.6	92.3	7.1	15.1	33.2	27.0
	SC	49.8	100.0	0.0	13.4	27.1	22.6	50.3	99.4	0.6	13.5	29.3	23.4
	ReAct	51.0	96.8	2.8	10.7	26.5	22.7	50.7	97.4	2.6	11.2	27.7	23.1
	Reflexion	50.2	99.2	0.8	13.4	28.8	23.3	50.3	99.5	0.5	13.6	29.2	23.4
	Inner Th.	51.0	96.4	3.2	15.6	31.0	25.2	50.8	97.4	2.5	14.6	31.3	24.8
	Proact.	**52.0**	**16.2**	**9.7**	**44.8**	**47.7**	**38.6**	**53.5**	**18.9**	**13.8**	**44.0**	**48.0**	**39.8**
	Reason.-high	**52.0**	91.5	7.7	16.8	31.8	27.1	52.0	93.0	6.2	15.2	32.3	26.4
Qwen2.5-7B-Inst.	Direct	**58.5**	**72.1**	**21.5**	**18.7**	**33.8**	**33.1**	54.8	77.9	15.0	**17.9**	**33.6**	30.3
	CoT	55.7	80.6	15.8	17.8	31.7	30.2	53.6	84.5	12.3	16.3	31.6	28.5
	SC	56.1	75.3	17.8	15.6	29.8	29.8	54.1	79.4	14.2	15.9	31.6	28.9
	ReAct	56.7	**72.1**	**21.5**	14.8	32.1	31.3	**55.4**	**74.7**	**18.1**	15.2	33.3	**30.5**
	Reflexion	55.7	83.8	15.0	14.3	30.1	28.8	52.6	89.6	8.9	13.3	29.9	26.2
	Inner Th.	50.4	99.2	0.8	13.5	27.4	23.0	50.2	99.4	0.6	12.5	27.8	22.8
	Proact.	50.4	97.2	2.0	14.6	26.1	23.3	50.5	97.0	2.3	15.3	26.4	23.6

Training Methods. Table 4 shows that executable supervision followed by the PACT reward gives the strongest POISE result on both trained backbones. On Qwen2.5-1.5B-Instruct, PACT adds 5.9 POISE points and 8.2 pair-consistency points over SFT. The larger pair-consistency gain suggests greater sensitivity to decisive condition changes. Outcome- and judge-reward updates yield smaller POISE gains, while GRPO initialized from the base model remains well below supervised training on both backbones. On CAB, score calibration and tool selection improve, although missed detection rises relative to SFT (Table 22). Together, the results support label-derived initialization and verifiable optimization with metric-dependent transfer.

5.2 FRAMEWORK ANALYSIS

Hyperparameter Sensitivity. Figure 4 shows that PACT performs best near the default setting across four optimization controls. POISE peaks at 89.0 with five rollouts per input, while both weak

Table 4: POISE and CAB results for two backbones. Bold and underlined values mark the best and second-best scores per backbone. Δ compares SFT+PACT with SFT.

		POISE								CAB			
Training	Method	POISE↑	Acc-P↑	Pair-C.↑	Tool-F1↑	Chan.-F1↑	DVI↑	MSVR↓	RMSE↓	Index↑	Score↑	Tool-F1↑	RMSE↓
Qwen2.5-1.5B-Instruct													
Prompting	Direct	19.5	49.7	1.2	2.6	24.6	10.0	3.9	1.600	54.5	2.3	45.5	3.117
	ProAgent	24.5	49.9	2.3	9.8	35.8	8.2	1.8	1.675	55.3	6.5	45.9	1.756
GRPO	Outcome	27.3	51.8	10.5	5.8	41.0	10.1	0.6	1.656	88.0	42.5	53.1	1.401
	PACT (Ours)	45.3	54.6	26.6	50.0	50.0	0.0	**0.0**	1.945	88.0	62.8	56.5	1.372
SFT	–	83.1	88.7	78.7	81.4	83.7	64.8	0.3	1.089	91.5	65.5	73.6	1.091
SFT + GRPO	Outcome	85.1	91.2	82.6	82.0	84.5	63.1	0.3	0.985	**93.3**	65.8	**75.2**	**1.039**
	Judge	86.7	91.8	83.6	85.4	86.1	65.3	0.2	0.941	91.5	**69.0**	74.9	1.135
	PACT (Ours)	**89.0**	**93.3**	**86.9**	**86.8**	**89.0**	**66.5**	0.1	**0.864**	92.3	65.8	74.7	1.063
Δ		+5.9	+4.6	+8.2	+5.4	+5.3	+1.7	−0.2	−0.225	+0.8	+0.3	+1.0	−0.028
Qwen3-1.7B													
Prompting	Direct	21.5	50.2	0.4	7.1	28.4	9.5	**0.2**	**1.557**	81.5	15.5	50.8	1.767
	ProAgent	23.7	51.1	7.1	10.8	25.7	11.4	0.8	1.612	63.3	24.0	45.4	2.237
GRPO	Outcome	23.3	51.4	4.6	9.1	28.2	10.2	**0.2**	1.653	82.5	30.5	47.5	1.61
	PACT (Ours)	22.5	49.9	1.3	12.3	26.4	8.9	1.0	2.030	80.5	54.0	53.7	1.53
SFT	–	42.1	56.1	18.3	45.6	48.4	31.4	1.0	1.912	83.5	54.3	55.1	1.544
SFT + GRPO	Outcome	44.0	56.4	26.9	43.6	49.1	24.5	4.1	2.136	**88.0**	54.8	58.7	**1.31**
	Judge	40.6	52.4	9.8	**49.3**	**50.9**	29.8	1.4	2.039	86.0	53.8	**60.5**	1.357
	PACT (Ours)	**46.6**	**58.2**	**28.6**	49.1	50.3	**32.4**	**0.2**	1.880	85.8	**61.0**	57.5	1.510
Δ		+4.5	+2.1	+10.3	+3.5	+1.9	+1.0	−0.8	−0.032	+2.3	+6.7	+2.4	−0.034

clipping and too few rollouts reduce pair consistency. Increasing the evidence weight improves DVI but gradually reduces the composite score beyond the default. These trends suggest that adequate sampling and bounded updates preserve useful pairwise reward contrast.

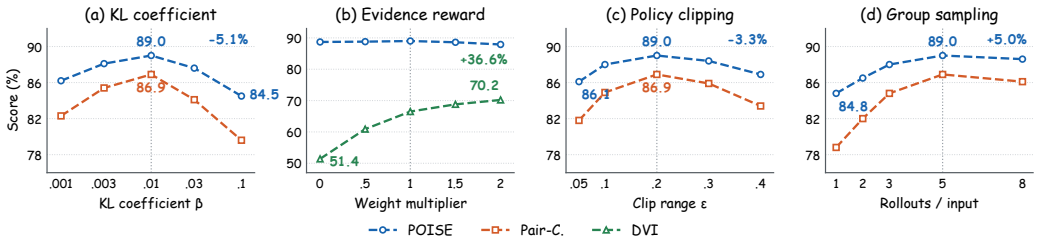

Figure 4: PACT sensitivity. Percentages compare labeled endpoints.

Ablation Studies. Table 5 shows that the reward components support distinct capabilities that the aggregate POISE score can obscure. Removing the evidence term reduces decisive-variable identification by 15.1 points, while removing safety raises the violation rate by 2.4 points. Safety and timing omissions barely affect the composite score despite their targeted effects. The safety term has a small weight and applies only to actionable outputs, so its omission can raise violations without a comparable shift in the mean score. Evidence removal instead weakens attribution to the rule's decisive variable. These distinct failure modes support evaluating evidence and delivery constraints alongside the aggregate POISE score in proactive deployments.

Table 5: Reward component ablation on Qwen2.5-1.5B-Instruct.

Reward	POISE↑	Acc-P↑	Pair-C.↑	Tool-F1↑	Chan.-F1↑	DVI↑	MSVR↓	Action.↑	RMSE↓
SFT, no RL	83.1	88.7	78.7	81.4	83.7	64.8	0.3	99.7	1.089
PACT reward	89.0	93.3	86.9	86.8	89.0	66.5	0.1	99.8	0.864
w/o decision	85.7	**88.1**	82.5	84.9	87.2	64.2	0.2	98.9	0.913
w/o evidence	88.7	93.2	86.7	86.4	88.6	**51.4**	0.2	99.7	0.866
w/o tool	85.3	91.4	85.2	**77.2**	87.5	65.0	0.1	99.7	0.888
w/o channel	87.0	92.6	86.3	85.8	**83.3**	65.8	0.1	99.7	0.872
w/o safety	88.8	93.1	86.8	86.5	88.9	66.3	**2.5**	99.8	0.865
w/o timing	88.9	93.2	86.8	86.6	88.9	66.1	0.1	**94.2**	0.866

Scaling Behavior. Fig. 5 shows that data coverage has a larger association with POISE performance than backbone size in the tested range. Expanding SFT data from 10% to 100% adds 37.0

POISE points, whereas scaling the fully supervised backbone from 1.5B to 7B adds 3.0 points. The PACT curve retains the sharp rise at full data while showing a moderate advantage over SFT across the smaller fractions. These trends are consistent with coverage of distinct decision conditions playing a larger role than parameter count in this setting.

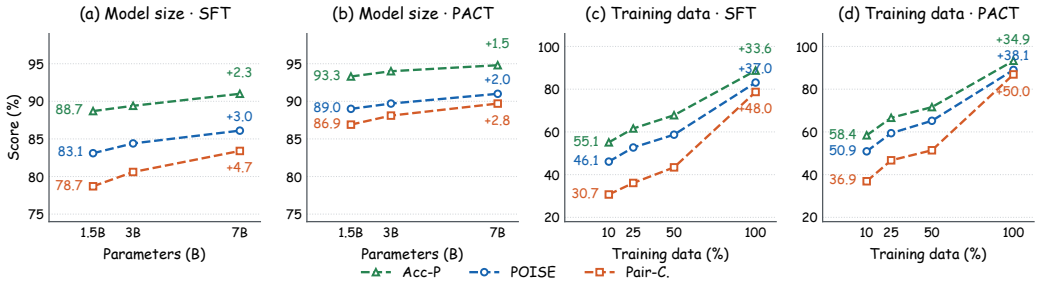

Figure 5: POISE scaling by model size and data. Labels show initial scores and gains.

Generalization. Table 6 shows uneven degradation when a scenario is omitted from training. For Airport flow, pair consistency falls from 83.8 to 11.9. Scenarios with recurring sensor patterns retain more of their in-domain performance than cases that depend on scenario-specific context, such as boarding state. Across scenarios, pair consistency declines more than decision accuracy when held out. This suggests that intervention behavior transfers more readily than the conditions governing when to act.

Table 6: Leave-one-scenario-out results. In/Out: training with-/without the tested scenario.

		POISE↑		Acc-P↑		Pair-C.↑		Chan.-F1↑	
Tier	Scenario	In	Out	In	Out	In	Out	In	Out
REFLEX	Rear approach	92.5	53.3	93.8	63.6	88.3	39.0	93.8	59.2
	Road crossing	57.6	41.1	66.9	52.6	41.2	16.2	61.1	49.4
INTERACTIVE	Elder routine	82.3	41.3	88.1	52.6	77.8	12.1	88.1	51.4
	Focus shield	82.0	55.8	86.6	69.9	74.0	47.1	82.8	56.6
	Heat stress	83.6	43.4	92.4	54.6	85.8	17.3	81.1	52.3
	Talk pacing	82.4	47.7	89.9	60.4	79.9	37.7	81.9	51.4
	Venue norms	80.9	41.6	85.7	53.0	71.9	17.3	83.0	49.1
	PA announcement	89.9	42.5	94.4	55.6	88.7	14.1	88.0	50.5
	Cross-language assist	87.0	48.8	89.6	61.8	79.2	31.9	89.6	52.7
DEFERRED	Airport flow	88.0	37.6	91.4	49.5	83.8	11.9	88.1	45.1
	Child safety	91.3	44.4	92.7	60.4	87.2	28.7	92.7	43.0
	Meeting prep	80.1	41.1	94.3	55.2	88.5	19.5	68.8	45.0

Case Study. Figure 6 shows SFT+PACT selecting silence, haptic, two-device, and staged responses across four settings. It avoids SFT's false airport intervention and Outcome's missed heat alert, while the child-safety case matches both delivery channels. The venue case matches tools and channels but differs from the reference in screen wording, separating delivery coverage from content fidelity. Further baseline cases appear in Appendix C.

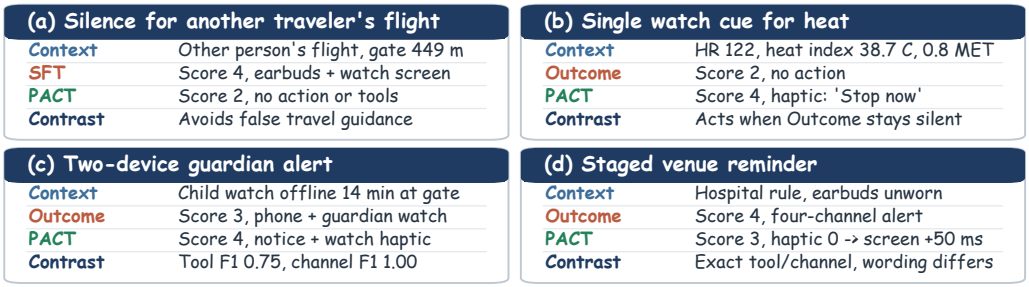

Figure 6: Qwen3-1.7B SFT+PACT responses with matched SFT/Outcome comparisons.

6 CONCLUSION

This paper proposes PACT, a training framework that turns executable rules into counterfactual supervision and verifiable rewards for proactive policy learning. We introduce POISE, a benchmark of 8559 wearable decisions that evaluates whether, what, and how an agent should intervene under sensing, delivery, and timing constraints. PACT outperforms supervised and RL objectives on POISE and improves tool selection and proactivity-score calibration on an external benchmark.

AI USE STATEMENT

In this work, we used generative AI tools to assist with grammar and manuscript wording and to check mathematical notation. Additionally, we used these tools to generate selected scenario illustrations in Figures 1 and 3. We also used an LLM to render POISE's natural-language contexts and fixed-label annotations from structured scenarios and recorded sensor or contextual inputs. We take responsibility for the final content, including AI-assisted text and figures. These tools did not independently determine the methods or technical conclusions.

REPRODUCIBILITY STATEMENT

The code and demo samples for POISE are available at `https://anonymous.4open.science/r/PACT-040B/`. Sections 3 and 5 specify PACT and the evaluation protocol. Appendix A documents POISE construction, data grounding, split statistics, and human validation. Appendix A.1 defines the metrics, Appendix B.1 details the baseline protocols, and Appendix B.3 gives the reward definitions, optimization objective, and training configuration. Appendices B–B.5 report full results and additional analyses, and Appendix C provides case studies.

REFERENCES

Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and context-awareness. In *International symposium on handheld and ubiquitous computing*, pp. 304–307. Springer, 1999.

DeepSeek-AI, Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenhao Xu, Chong Ruan, Damai Dai, Daya Guo, Dejian Yang, Deli Chen, Erhang Li, Fangqi Zhou, Fangyun Lin, Fucong Dai, Guangbo Hao, Guanting Chen, Guowei Li, H. Zhang, Hanwei Xu, Hao Li, Haofen Liang, Haoran Wei, Haowei Zhang, Haowen Luo, Haozhe Ji, Honghui Ding, Hongxuan Tang, Huanqi Cao, Huazuo Gao, Hui Qu, Hui Zeng, Jialiang Huang, Jiashi Li, Jiaxin Xu, Jiewen Hu, Jingchang Chen, Jingting Xiang, Jingyang Yuan, Jingyuan Cheng, Jinhua Zhu, Jun Ran, Junguang Jiang, Junjie Qiu, Junlong Li, Junxiao Song, Kai Dong, Kaige Gao, Kang Guan, Kexin Huang, Kexing Zhou, Kezhao Huang, Kuai Yu, Lean Wang, Lecong Zhang, Lei Wang, Liang Zhao, Liangsheng Yin, Lihua Guo, Lingxiao Luo, Linwang Ma, Litong Wang, Liyue Zhang, M. S. Di, M. Y Xu, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Mingxu Zhou, Panpan Huang, Peixin Cong, Peiyi Wang, Qiancheng Wang, Qihao Zhu, Qingyang Li, Qinyu Chen, Qiushi Du, Ruiling Xu, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, Runqiu Yin, Runxin Xu, Ruomeng Shen, Ruoyu Zhang, S. H. Liu, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shaofei Cai, Shaoyuan Chen, Shengding Hu, Shengyu Liu, Shiqiang Hu, Shirong Ma, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan, Songyang Zhou, Tao Ni, Tao Yun, Tian Pei, Tian Ye, Tianyuan Yue, Wangding Zeng, Wen Liu, Wenfeng Liang, Wenjie Pang, Wenjing Luo, Wenjun Gao, Wentao Zhang, Xi Gao, Xiangwen Wang, Xiao Bi, Xiaodong Liu, Xiaohan Wang, Xiaokang Chen, Xiaokang Zhang, Xiaotao Nie, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xingkai Yu, Xingyou Li, Xinyu Yang, Xinyuan Li, Xu Chen, Xuecheng Su, Xuehai Pan, Xuheng Lin, Xuwei Fu, Y. Q. Wang, Yang Zhang, Yanhong Xu, Yanru Ma, Yao Li, Yao Li, Yao Zhao, Yaofeng Sun, Yaohui Wang, Yi Qian, Yi Yu, Yichao Zhang, Yifan Ding, Yifan Shi, Yiliang Xiong, Ying He, Ying Zhou, Yinmin Zhong, Yishi Piao, Yisong Wang, Yixiao Chen, Yixuan Tan, Yixuan Wei, Yiyang Ma, Yiyuan Liu, Yonglun Yang, Yongqiang Guo, Yongtong Wu, Yu Wu, Yuan Cheng, Yuan Ou, Yuanfan Xu, Yuduan Wang, Yue Gong, Yuhan Wu, Yuheng Zou, Yukun Li, Yunfan Xiong, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuyang Zhou, Z. F. Wu, Z. Z. Ren, Zehua Zhao, Zehui Ren, Zhangli Sha, Zhe Fu, Zhean Xu, Zhenda Xie, Zhengyan Zhang, Zhewen Hao, Zhibin Gou, Zhicheng Ma, Zhigang Yan, Zhihong Shao, Zhixian Huang, Zhiyu Wu, Zhuoshu Li, Zhuping Zhang, Zian Xu, Zihao Wang, Zihui Gu, Zijia Zhu, Zilin Li, Zipeng Zhang, Ziwei Xie, Ziyi Gao, Zizheng Pan, Zongqing Yao, Bei Feng, Hui Li, J. L. Cai, Jiaqi Ni, Lei Xu, Meng Li, Ning Tian, R. J. Chen, R. L. Jin, S. S. Li, Shuang Zhou, Tianyu Sun, X. Q. Li, Xiangyue Jin, Xiaojin Shen, Xiaosha Chen, Xinnan Song, Xinyi Zhou, Y. X. Zhu, Yanping Huang, Yaohui Li, Yi Zheng, Yuchen Zhu, Yunxian Ma, Zhen Huang, Zhipeng Xu, Zhongyu Zhang, Dongjie Ji, Jian Liang, Jianzhong Guo, Jin Chen, Leyi Xia, Miaojun Wang, Mingming Li, Peng Zhang, Ruyi Chen, Shangmian Sun, Shaoqing Wu, Shengfeng Ye, T. Wang, W. L. Xiao,

Wei An, Xianzu Wang, Xiaowen Sun, Xiaoxiang Wang, Ying Tang, Yukun Zha, Zekai Zhang, Zhe Ju, Zhen Zhang, and Zihua Qu. Deepseek-v3.2: Pushing the frontier of open large language models, 2025. URL https://arxiv.org/abs/2512.02556.

Yang Deng, Wenqiang Lei, Wai Lam, and Tat-Seng Chua. A survey on proactive dialogue systems: Problems, methods, and prospects. *arXiv preprint arXiv:2305.02750*, 2023.

Anind K Dey. Understanding and using context. *Personal and ubiquitous computing*, 5(1):4–7, 2001.

Leo Gao, John Schulman, and Jacob Hilton. Scaling laws for reward model overoptimization. In *International conference on machine learning*, pp. 10835–10866. PMLR, 2023.

Matt Gardner, Yoav Artzi, Victoria Basmov, Jonathan Berant, Ben Bogin, Sihao Chen, Pradeep Dasigi, Dheeru Dua, Yanai Elazar, Ananth Gottumukkala, et al. Evaluating models' local decision boundaries via contrast sets. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 1307–1323, 2020.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, Bing Xue, Bingxuan Wang, Bochao Wu, Bei Feng, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chong Ruan, Damai Dai, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao, Guanting Chen, Guowei Li, H. Zhang, Hanwei Xu, Honghui Ding, Huazuo Gao, Hui Qu, Hui Li, Jianzhong Guo, Jiashi Li, Jingchang Chen, Jingyang Yuan, Jinhao Tu, Junjie Qiu, Junlong Li, J. L. Cai, Jiaqi Ni, Jian Liang, Jin Chen, Kai Dong, Kai Hu, Kaichao You, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang, Liang Zhao, Litong Wang, Liyue Zhang, Lei Xu, Leyi Xia, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Mingxu Zhou, Meng Li, Miaojun Wang, Mingming Li, Ning Tian, Panpan Huang, Peng Zhang, Qiancheng Wang, Qinyu Chen, Qiushi Du, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, R. J. Chen, R. L. Jin, Ruyi Chen, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shengfeng Ye, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan, S. S. Li, Shuang Zhou, Shaoqing Wu, Tao Yun, Tian Pei, Tianyu Sun, T. Wang, Wangding Zeng, Wen Liu, Wenfeng Liang, Wenjun Gao, Wenqin Yu, Wentao Zhang, W. L. Xiao, Wei An, Xiaodong Liu, Xiaohan Wang, Xiaokang Chen, Xiaotao Nie, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xinyu Yang, Xinyuan Li, Xuecheng Su, Xuheng Lin, X. Q. Li, Xiangyue Jin, Xiaojin Shen, Xiaosha Chen, Xiaowen Sun, Xiaoxiang Wang, Xinnan Song, Xinyi Zhou, Xianzu Wang, Xinxia Shan, Y. K. Li, Y. Q. Wang, Y. X. Wei, Yang Zhang, Yanhong Xu, Yao Li, Yao Zhao, Yaofeng Sun, Yaohui Wang, Yi Yu, Yichao Zhang, Yifan Shi, Yiliang Xiong, Ying He, Yishi Piao, Yisong Wang, Yixuan Tan, Yiyang Ma, Yiyuan Liu, Yongqiang Guo, Yuan Ou, Yuduan Wang, Yue Gong, Yuheng Zou, Yujia He, Yunfan Xiong, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuyang Zhou, Y. X. Zhu, Yanping Huang, Yaohui Li, Yi Zheng, Yuchen Zhu, Yunxian Ma, Ying Tang, Yukun Zha, Yuting Yan, Z. Z. Ren, Zehui Ren, Zhangli Sha, Zhe Fu, Zhean Xu, Zhenda Xie, Zhengyan Zhang, Zhewen Hao, Zhicheng Ma, Zhigang Yan, Zhiyu Wu, Zihui Gu, Zijia Zhu, Zijun Liu, Zilin Li, Ziwei Xie, Ziyang Song, Zizheng Pan, Zhen Huang, Zhipeng Xu, Zhongyu Zhang, and Zhen Zhang. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645 (8081):633–638, September 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-09422-z. URL http://dx.doi.org/10.1038/s41586-025-09422-z.

Eric Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 159–166, 1999.

Divyansh Kaushik, Eduard Hovy, and Zachary C Lipton. Learning the difference that makes a difference with counterfactually-augmented data. *arXiv preprint arXiv:1909.12434*, 2019.

Yubin Kim, Xuhai Xu, Daniel McDuff, Cynthia Breazeal, and Hae Won Park. Health-llm: Large language models for health prediction via wearable sensor data. *arXiv preprint arXiv:2401.06866*, 2024.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model

serving with pagedattention. In *Proceedings of the 29th symposium on operating systems principles*, pp. 611–626, 2023.

Chenyi Li, Guande Wu, Gromit Yeuk-Yin Chan, Dishita Gdi Turakhia, Sonia Castelo Quispe, Dong Li, Leslie Welch, Claudio Silva, and Jing Qian. Satori: Towards proactive ar assistant with belief-desire-intention user modeling. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–24, 2025.

Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. In *International Conference on Learning Representations*, volume 2024, pp. 39578–39601, 2024.

Xingyu Bruce Liu, Shitao Fang, Weiyan Shi, Chien-Sheng Wu, Takeo Igarashi, and Xiang'Anthony' Chen. Proactive conversational agents with inner thoughts. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–19, 2025.

Yaxi Lu, Shenzhi Yang, Cheng Qian, Guirong Chen, Qinyu Luo, Yesai Wu, Huadong Wang, Xin Cong, Zhong Zhang, Yankai Lin, et al. Proactive agent: Shifting llm agents from reactive responses to active assistance. In *International Conference on Learning Representations*, volume 2025, pp. 47431–47457, 2025.

Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. My phone and me: understanding people's receptivity to mobile notifications. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 1021–1032, 2016.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35: 27730–27744, 2022.

Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. Librispeech: an asr corpus based on public domain audio books. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 5206–5210. IEEE, 2015.

Qwen, :, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report, 2025. URL https://arxiv.org/abs/2412.15115.

Attila Reiss and Didier Stricker. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th international symposium on wearable computers*, pp. 108–109. IEEE, 2012.

Attila Reiss, Ina Indlekofer, Philip Schmidt, and Kristof Van Laerhoven. Deep ppg: Large-scale heart rate estimation with convolutional neural networks. *Sensors*, 19(14):3079, 2019.

Marco Tulio Ribeiro, Tongshuang Wu, Carlos Guestrin, and Sameer Singh. Beyond accuracy: Behavioral testing of nlp models with checklist. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pp. 4902–4912, 2020.

Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. *Advances in neural information processing systems*, 36:68539–68551, 2023.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017. URL https://arxiv.org/abs/1707.06347.

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, 2024. URL https://arxiv.org/abs/2402.03300.

Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. Hybridflow: A flexible and efficient rlhf framework. In *Proceedings of the Twentieth European Conference on Computer Systems*, pp. 1279–1297, 2025.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. *Advances in neural information processing systems*, 36:8634–8652, 2023.

Joaquín Torres-Sospedra, Raúl Montoliu, Adolfo Martínez-Usó, Joan P Avariento, Tomás J Arnau, Mauri Benedito-Bordonau, and Joaquín Huerta. Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems. In *2014 international conference on indoor positioning and indoor navigation (IPIN)*, pp. 261–270. IEEE, 2014.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models, 2023. URL https://arxiv.org/abs/2203.11171.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

Huatao Xu, Liying Han, Qirui Yang, Mo Li, and Mani Srivastava. Penetrative ai: Making llms comprehend the physical world. In *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications*, pp. 1–7, 2024.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report, 2025. URL https://arxiv.org/abs/2505.09388.

Bufang Yang, Lilin Xu, Liekang Zeng, Kaiwei Liu, Siyang Jiang, Wenrui Lu, Hongkai Chen, Xiaofan Jiang, Guoliang Xing, and Zhenyu Yan. Contextagent: Context-aware proactive llm agents with open-world sensory perceptions. *Advances in Neural Information Processing Systems*, 38: 167509–167543, 2026.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, YuYue, Weinan Dai, Tiantian Fan, Gaohong Liu, Juncai Liu, LingJun Liu, Xin Liu, Haibin Lin, Zhiqi Lin, Bole Ma, Guangming Sheng, Yuxuan Tong, Chi Zhang, Mofan Zhang, Ru Zhang, Wang Zhang, Hang Zhu, Jinhua Zhu, Jiaze Chen, Jiangjie Chen, Chengyi Wang, Hongli Yu, Yuxuan Song, Xiangpeng Wei, Hao Zhou, Jingjing Liu, Wei-Ying Ma, Ya-Qin Zhang, Lin Yan, Yonghui Wu, and Mingxuan Wang. DAPO: An open-source LLM reinforcement learning system at scale. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=2a36EMSSTp.

Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, Zheyan Luo, Zhangchi Feng, and Yongqiang Ma. Llamafactory: Unified efficient fine-tuning of 100+ language models, 2024. URL https://arxiv.org/abs/2403.13372.

Yu Zheng, Xing Xie, Wei-Ying Ma, et al. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, 33(2):32–39, 2010.

Roadmap. Appendix A documents POISE construction, text rendering, data grounding, and human validation. Appendix A.1 defines the metrics. Appendices B–B.5 give baseline protocols, full results, reward and training details, and subgroup analyses. Appendix C presents case studies.

A THE POISE BENCHMARK

Construction Pipeline. Each sample follows its scenario's executable specification, whose predicates fix the label (Section 3). *Sampling* draws scenario variables from the specification and recorded sensor or contextual inputs, retaining counterfactuals only when the label flips. *Enrichment* computes decisive variables, actionable windows, evidence confidence and abstention status without a model. For *rendering*, sampled values fill a scenario template to form a factual skeleton. GPT-5.4 receives this skeleton, the signal snapshot, device availability and the truth values of trigger and suppression facts, then writes the fused context and a coarser on-device perception report. Its prompt requires the numeric values in the skeleton and predicate-relevant facts to remain recoverable, while excluding unsupported details and intervention judgments. Given the fixed label, the same model generates evidence, tools and delivery plans during *annotation* without deciding whether to intervene. Deterministic *repair* normalizes unambiguous fields, and *personas* retain sensitive-attribute provenance. *Blind critics* flag disagreements with the predicates, while *validation* applies nine checks to schema, tools, channels, latency, pair integrity, language, consent, escalation and vocabulary leakage.

Dataset Statistics. Table 7 lists the twelve scenarios, their data sources, and the number of samples backed by the cited datasets. Table 8 compares POISE with CAB and CAB-Lite. For POISE, level 1 holds the non-intervening samples and levels 2 and 3 the intervening ones at increasing proactivity score, while CAB levels follow the CAB release. Latency tiers are reflex, interactive and deferred. Length is the mean and maximum word count, of the full prompt for POISE and of the sensory text for CAB. In Fig. 8, scores of three and above are proactive, and a proactive decision calls at least two tools, one to sense and one to act, while a non-proactive decision calls none. Table 9 lists, for each condition grounded in a cited dataset, the statistic measured in the recording and the sampling range or threshold it fixes, and Fig. 7 shows the windows of those recordings on which the sensor-side conditions of a trigger hold. Figure 9 shows representative raw signals from the five cited datasets using spectral, spatial, and Wi-Fi views appropriate to each modality.

Table 7: The twelve POISE scenarios. *Public* counts samples using one of five public recordings. The three scenarios with zero public samples use author-recorded data.

Scenario	Short	Dominant tier	n	Public	Source
Rear approach	Rear	reflex	816	336	PAMAP2
Road crossing	Cross	reflex	680	280	PAMAP2
Elder routine	Elder	interactive	952	392	PAMAP2
Focus shield	Focus	interactive	535	224	PAMAP2
Heat stress	Heat	interactive	952	392	PPG-DaLiA
Talk pacing	Talk	interactive	816	336	LibriSpeech
Venue norms	Venue	interactive	952	392	GeoLife
PA announcement	Gate	interactive	400	0	Author-recorded
Cross-language assist	Lang	interactive	400	0	Author-recorded
Airport flow	Airport	deferred	948	388	UJIIndoorLoc
Child safety	Child	deferred	708	336	GeoLife
Meeting prep	Meet	deferred	400	0	Author-recorded
Total			8,559	3,076	

Human Validation. Three annotators from the research group, all with experience in wearable sensing, rate ten complete counterfactual pairs from each of the twelve scenarios, covering 240 individual samples. They received no separate compensation for this evaluation. Table 10 gives the rubric verbatim. For the first three dimensions, each annotator sees one rendered situation, its persona and delivery constraints, but neither the derived label nor the recorded decisive variable nor



Figure 7: The POISE wordmark, drawn from three-second windows of the public recordings behind the benchmark. Red traces are windows on which the sensor-side conditions of a trigger hold, at their natural rate.

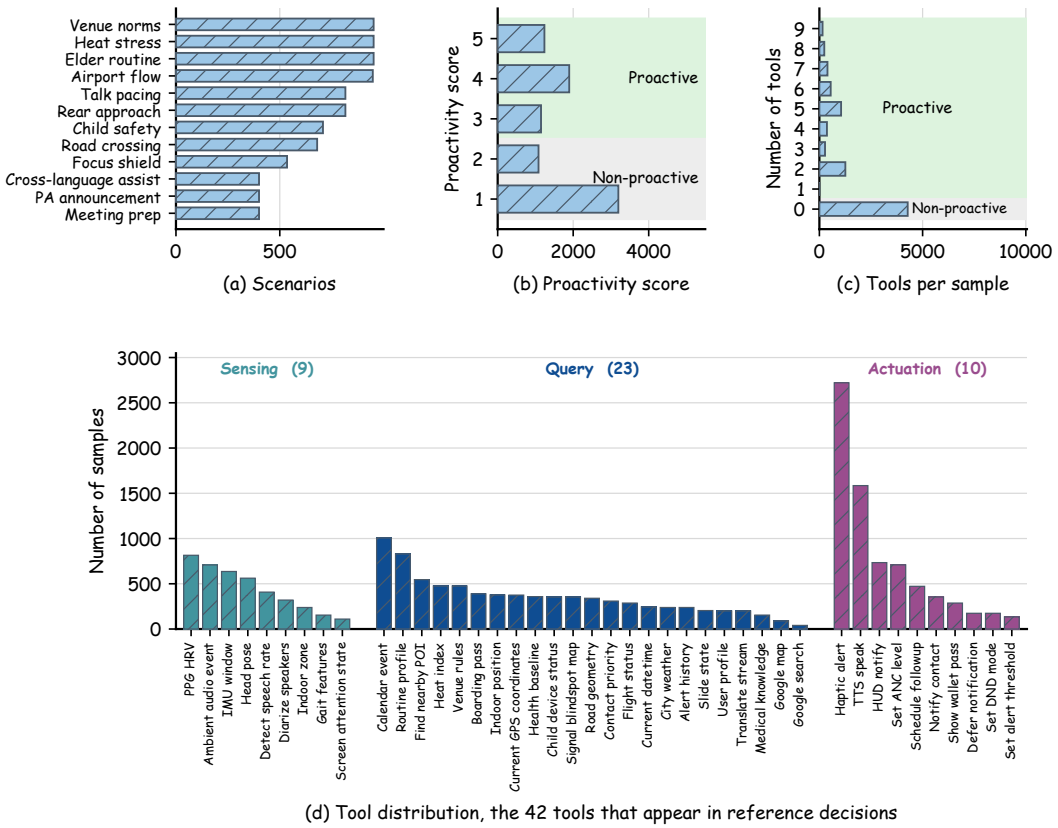

Figure 8: Statistics of POISE: (a) samples per scenario, (b) proactivity score, (c) tool calls per sample, and (d) frequency of the 42 tools in reference decisions, grouped by layer.

the other half of its pair. Decision agreement and decisive-variable attribution thus measure whether a reader independently arrives at the executable label. After these individual ratings are locked, the annotators see both halves to rate the separability of each pair. Delivery feasibility is rated on the 120 nonempty reference plans from the intervening halves. Every item is rated by all three annotators before any disagreement is discussed. Table 11 reports the results.

Table 8: Statistics of POISE, CAB and CAB-Lite.

	POISE		POISE-Lite	CAB		CAB-Lite	
	Train	Test	Test	Train	Test	Train	Test
Samples	5,134	3,425	494	600	400	180	120
Scenarios / categories	12 / 7	12 / 7	12 / 7	— / 9	— / 9	— / 9	— / 9
Act : silent	0.500	0.500	0.500	0.480	0.502	0.539	0.408
Counterfactual pairs	2,569	1,713	247	–	–	–	–
Levels	2,566/777/1,791	1,712/481/1,232	247/91/156	–	–	–	–
Latency tiers	916/3,034/1,184	580/1,973/872	78/312/104	–	–	–	–
Real-sensor samples	1,792	1,284	164	–	–	–	–
Signal variables / sample	20.9	21.1	20.7	–	–	–	–
Distinct tools	42	42	42	19	20	16	16
Samples with persona	3,919	2,474	342	572	383	169	117
Length	2,557/2,667	2,558/2,658	2,555/2,650	54/129	55/127	42/119	41/95

Table 9: Calibration of scenario conditions against the five public recording corpora.

Corpus	Condition	Statistic in the recordings	Resulting rule
PAMAP2	Gait stability	Autocorrelation 0.1 standing, 0.77 walking	Gate on ambulation
PPG-DaLiA	Sedentary, elevated heart rate	353 of 4,710 windows	Feasible region set accordingly
LibriSpeech	Talk pacing deviation	19.8 % at the 90th percentile	Threshold at 20 %
UJIIndoorLoc	Indoor fix quality	Median 5.5 m, wrong floor in 10.2 % of windows	Sampling range and the 25 m usability gate
GeoLife	Heading toward a venue	68° GPS jitter	Heading not used as a trigger
GeoLife	Child route deviation	Normal travel median 501 m	Threshold set above normal travel

A.1 Evaluation Metrics

This section defines every metric reported in the paper. The definitions follow the released scorer, which is shared by the zero-shot benchmark and the training pipeline.

Notation. The test set has N samples indexed by i. Sample i carries a reference proactivity score $s_i \in \{1, \dots, 5\}$, a reference tool set T_i, a reference delivery plan Π_i and the situation fields used below. The agent returns a score $\hat{s}_i$, a tool set $\hat{T}_i$ and a plan $\hat{\Pi}_i$. A response with no parseable decision object after three attempts is scored as $\hat{s}_i = 0$ with empty tools and plan, so it counts as staying silent. With threshold $\theta = 3$, the reference and predicted decisions are

$$y_i = \mathbb{1}[s_i \geq \theta], \qquad \hat{y}_i = \mathbb{1}[\hat{s}_i \geq \theta], \tag{5}$$

and $\mathcal{P} = \{i : y_i = 1\}$ and $\mathcal{N} = \{i : y_i = 0\}$ are the samples that should and should not trigger an intervention. For a predicted set $\hat{A}$ and a reference set A,

$$\mathrm{P}(\hat{A}, A) = \frac{|\hat{A} \cap A|}{|\hat{A}|}, \qquad \mathrm{R}(\hat{A}, A) = \frac{|\hat{A} \cap A|}{|A|}, \qquad \mathrm{F}_1(\hat{A}, A) = \frac{2\,\mathrm{P}\,\mathrm{R}}{\mathrm{P} + \mathrm{R}}, \tag{6}$$

with $\mathrm{P} = \mathrm{R} = \mathrm{F}_1 = 1$ when both sets are empty, a ratio set to 0 when only its denominator set is empty, and $\mathrm{F}_1 = 0$ when $\mathrm{P} + \mathrm{R} = 0$. Several metrics are defined only on a subset $\mathcal{I} \subseteq \{1, \dots, N\}$ of samples, stated for each metric below. A metric whose subset is empty is reported as a dash.

CAB-Comparable Decision and Tool Metrics.

$$\text{Acc-P} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}[\hat{y}_i = y_i], \qquad \text{RMSE} = \Big(\frac{1}{N} \sum_{i=1}^{N} (\hat{s}_i - s_i)^2 \Big)^{1/2}, \tag{7}$$

$$\text{MD} = \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} \mathbb{1}[\hat{y}_i = 0], \qquad \text{FD} = \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} \mathbb{1}[\hat{y}_i = 1], \tag{8}$$

$$\text{Tool-}X = \frac{1}{N} \sum_{i=1}^{N} X(\hat{T}_i, T_i), \qquad X \in \{\mathrm{P}, \mathrm{R}, \mathrm{F}_1\}, \tag{9}$$

where tool sets compare tool names only. The share of unparseable responses is Unp. $= \frac{1}{N} \sum_i u_i$, with $u_i = 1$ when sample i returned no decision object.

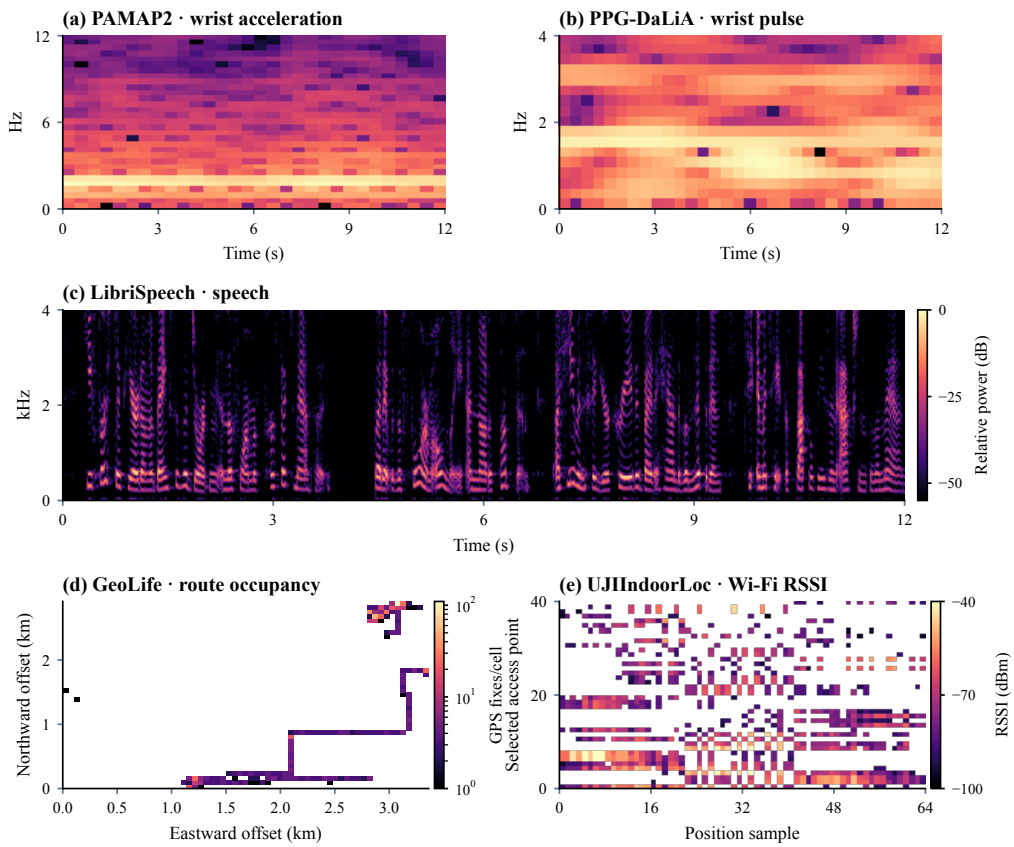

Figure 9: Representative recordings from five POISE source datasets. Panels (a)–(c) show independently normalized spectral power, (d) shows GPS fix density along one route, and (e) shows Wi-Fi RSSI on one floor. White Wi-Fi cells denote undetected access points.

Pair Consistency. Let $\mathcal{Q}$ be the set of counterfactual pairs (i, j) with both halves in the evaluated set. A pair counts only when both of its decisions are correct,

$$\text{Pair-C.} = \frac{1}{|\mathcal{Q}|} \sum_{(i,j)\in\mathcal{Q}} \mathbb{1}[\hat{y}_i = y_i]\mathbb{1}[\hat{y}_j = y_j]. \tag{10}$$

Because the two halves of a pair carry opposite labels, a policy that ignores the input decides at most one half correctly and scores 0. In a breakdown by level, scenario or latency tier, each pair is assigned to the group of its intervening half, and in a breakdown by suppression predicate to the group of its silent half.

Delivery Channels. A plan is a set of actions a, each with a device $d(a)$, a modality $m(a)$, a time offset $\Delta t(a)$ in milliseconds and a latency tier. Its channel set is $C(\Pi) = \{(d(a), m(a)) : a \in \Pi\}$, and

$$\text{Chan.-}X = \frac{1}{N} \sum_{i=1}^{N} X\big(C(\hat{\Pi}_i), C(\Pi_i)\big), \qquad X \in \{\mathrm{P}, \mathrm{R}, \mathrm{F}_1\}. \tag{11}$$

Safety. Sample i specifies the set of devices that are worn or carried, $\mathcal{W}_i$, a set of forbidden channels $\mathcal{F}_i$, and two indicators of situational constraints: $\kappa_i^{\text{vis}} = 1$ when the wearer must not shift their gaze, as when cycling or crossing a road, and $\kappa_i^{\text{aud}} = 1$ when the agent must not make a sound, as during a talk. Let M_{vis} be the modalities that require looking at a display, namely on-screen and head-up text, and M_{aud} the modalities that produce sound, namely speech and audio. An action is unsafe when

$$\begin{aligned} v_i(a) = \mathbb{1}\big[&d(a) \notin \mathcal{W}_i \;\vee\; (d(a), m(a)) \in \mathcal{F}_i \\ &\vee\; (\kappa_i^{\text{vis}} = 1 \wedge m(a) \in M_{\text{vis}}) \;\vee\; (\kappa_i^{\text{aud}} = 1 \wedge m(a) \in M_{\text{aud}})\big]. \end{aligned} \tag{12}$$

Table 10: Rubric used by the annotators of POISE. Each annotator sees the quoted wording verbatim.

NARRATIVE FAITHFULNESS — sample, binary

"Does the situation as written state every signal the decision predicates read, at the value the specification recorded?"

POSITIVE: Every predicate input appears with the same magnitude and sign, and nothing unmeasured is asserted as observed.

FAILURE: An input is missing, is paraphrased across its threshold, or the prose adds a condition the specification never sampled.

DECISION AGREEMENT — sample, binary

"Reading this as the user's wearable assistant, would you intervene here, and at what strength on the one-to-five proactivity scale?"

POSITIVE: The intervention side matches the derived label and the score is within one level of it.

FAILURE: The annotator takes the opposite side, or the score differs by two or more levels.

DECISIVE-VARIABLE ATTRIBUTION — sample, three-point ordinal

"Which single variable, had it taken a different value, would have reversed your decision?"

POSITIVE: *Exact* names the variable the enrichment stage recorded. *Partial* names a variable the predicate reads but not the one that flips.

FAILURE: *Wrong* names a correlated variable no predicate reads, or no variable can be named at all.

PAIR SEPARABILITY — counterfactual pair, binary

"These two situations were built to differ in one variable. Does that difference alone account for their opposite decisions?"

POSITIVE: The named variable is the only visible difference between the halves, and it is sufficient to reverse the decision.

FAILURE: The halves differ in more than the named variable, or both read as the same decision.

DELIVERY FEASIBILITY — delivery plan, five-point Likert

"Can the reference plan be carried out on the listed devices, inside the actionable window, and within the persona's consent state?"

POSITIVE: *5* every channel is available, permitted and timely, *4* one redundant or superfluous channel, *3* workable but the modality suits the situation poorly.

FAILURE: *2* the plan uses an occupied or absent channel or lands outside the window, *1* it records or addresses a third party who has not consented.

Table 11: Annotator ratings on the stratified sample. Rated: share of ratings at the positive levels of Table 10. α: Krippendorff's alpha.

Dimension	Unit	Items	Rated↑	α ↑
Narrative faithfulness	sample	240	96.0	0.82
Decision agreement	sample	240	88.6	0.73
Decisive-variable attribution	sample	240	82.1	0.68
Pair separability	pair	120	87.8	0.77
Delivery feasibility	delivery plan	120	93.1	0.79

The multi-channel safety-violation rate counts samples with at least one unsafe action,

$$\mathrm{MSVR} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\Big[\textstyle\sum_{a\in\hat{\Pi}_i} v_i(a) > 0\Big]. \tag{13}$$

Compliance violations are defined over the predicted tool calls. Let $\mathcal{T}_{\mathrm{rec}}$ be the tools that start a recording and $\mathcal{T}_{\mathrm{sos}}$ the tools that place an emergency call, let $g_i = 1$ when third-party consent to

recording is granted in sample i, and let w_t be the cancellation window in seconds that an emergency call t specifies. A call is non-compliant when it records without consent or places an emergency call without a cancellation window between 10 and 30 seconds,

$$\text{Compl.} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\Big[\exists\, t \in \hat{T}_i :\ \big(t \in \mathcal{T}_{\text{rec}} \wedge g_i = 0\big)\ \vee\ \big(t \in \mathcal{T}_{\text{sos}} \wedge w_t \notin [10, 30]\big)\Big], \tag{14}$$

where a missing window counts as outside the interval.

Ordering. Some situations require one channel to act before another, for example ramping noise cancellation down before speech. Let $\mathcal{R}_i$ be the required orderings (c_b, c_a) and $\hat{t}_i(c) = \min\{\Delta t(a) : a \in \hat{\Pi}_i,\ (d(a), m(a)) = c\}$ the earliest time the plan uses channel c. On $\mathcal{I}_{\text{ord}} = \{i : \mathcal{R}_i \text{ defined}\}$,

$$\text{Order.} = \frac{1}{|\mathcal{I}_{\text{ord}}|}\sum_{i \in \mathcal{I}_{\text{ord}}} \mathbb{1}\big[\exists\, (c_b, c_a) \in \mathcal{R}_i :\ c_b, c_a \in C(\hat{\Pi}_i)\ \wedge\ \hat{t}_i(c_b) \geq \hat{t}_i(c_a)\big]. \tag{15}$$

Timing. Sample i may carry an actionable window τ_i in seconds. The latency of a tool chain is the sum of per-tool costs $\ell(t)$, set by tool layer to 50 ms for sensing, 800 ms for query and 20 ms for actuation, and to 800 ms for an unknown tool. Let $r(a) = 1$ when action a belongs to the reflex tier. A reflex action does not wait for the tool chain, so the effective delivery time of an action is

$$e_i(a) = \Delta t(a) + \big(1 - r(a)\big)\sum_{t \in \hat{T}_i} \ell(t), \tag{16}$$

On $\mathcal{I}_{\text{act}} = \{i : \tau_i > 0,\ \hat{\Pi}_i \neq \emptyset\}$, a plan is actionable when its earliest delivery lands inside the window, and the relative deadline error compares the agent's stated deadline $\hat{\tau}_i$ with the window on $\mathcal{I}_{\text{dl}} = \{i : \tau_i > 0,\ \hat{\tau}_i \text{ stated}\}$,

$$\text{Action.} = \frac{1}{|\mathcal{I}_{\text{act}}|}\sum_{i \in \mathcal{I}_{\text{act}}} \mathbb{1}\Big[\min_{a \in \hat{\Pi}_i} e_i(a) \leq 1000\,\tau_i\Big], \qquad \text{Deadl.} = \frac{1}{|\mathcal{I}_{\text{dl}}|}\sum_{i \in \mathcal{I}_{\text{dl}}} \frac{|\hat{\tau}_i - \tau_i|}{\tau_i}. \tag{17}$$

Deadl. is reported for runs that state a deadline on at least 5 % of the evaluated samples, and a dash marks the rest.

Evidence. Sample i may name its decisive variables $V_i^\star$, the variables whose change flips the label across its counterfactual pair. On $\mathcal{I}_{\text{dvi}} = \{i : V_i^\star \neq \emptyset,\ \hat{v}_i \text{ stated}\}$, decisive-variable identification checks the variable the agent names,

$$\text{DVI} = \frac{1}{|\mathcal{I}_{\text{dvi}}|}\sum_{i \in \mathcal{I}_{\text{dvi}}} \mathbb{1}[\hat{v}_i \in V_i^\star]. \tag{18}$$

Where the executable specification grounds the reliability of the evidence as $c_i \in [0, 1]$ and the agent states a confidence $\hat{c}_i$, the confidence gap on $\mathcal{I}_{\text{conf}}$ is $\text{Conf.} = \frac{1}{|\mathcal{I}_{\text{conf}}|}\sum_{i \in \mathcal{I}_{\text{conf}}} |\hat{c}_i - c_i|$.

Abstention and Escalation. Let α_i indicate that abstaining is appropriate for sample i and $\hat{\alpha}_i$ that the agent abstained, on the samples $\mathcal{I}_{\text{abs}}$ where the agent reports an abstention decision. Then

$$\text{Abst.-P} = \frac{\sum_{i \in \mathcal{I}_{\text{abs}}} \alpha_i \hat{\alpha}_i}{\sum_{i \in \mathcal{I}_{\text{abs}}} \hat{\alpha}_i}, \qquad \text{Abst.-R} = \frac{\sum_{i \in \mathcal{I}_{\text{abs}}} \alpha_i \hat{\alpha}_i}{\sum_{i \in \mathcal{I}_{\text{abs}}} \alpha_i}. \tag{19}$$

An escalation ladder E reaches outside help, $\rho(E) = 1$, when one of its steps notifies a contact or calls emergency services. Over-escalation counts help summoned that the reference does not summon,

$$\text{Over-esc.} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\big[\rho(\hat{E}_i) = 1 \wedge \rho(E_i) = 0\big]. \tag{20}$$

Composite Score and Cost. The POISE score averages the four headline metrics,

$$\text{POISE} = \tfrac{1}{4}\big(\text{Acc-P} + \text{Pair-C.} + \text{Tool-F1} + \text{Chan.-F1}\big), \tag{21}$$

and is left undefined when any of the four is undefined. Cost is reported per sample as model calls (Calls), prompt plus completion tokens (Tok.) and wall-clock seconds (Sec.).

CAB and CAB-Lite Metrics. CAB and CAB-Lite are scored with the reference CAB script, reported in percent. With the same thresholded decisions,

$$\text{Index} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}[\hat{y}_i = y_i], \qquad \text{Score} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}[\hat{s}_i = s_i], \tag{22}$$

$$\text{MD}_{\text{CAB}} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}[y_i = 1 \wedge \hat{y}_i = 0], \qquad \text{FD}_{\text{CAB}} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}[y_i = 0 \wedge \hat{y}_i = 1]. \tag{23}$$

Two conventions differ from the POISE scorer. The CAB miss and false-detection rates are normalized by all N samples rather than by $|\mathcal{P}|$ and $|\mathcal{N}|$, and CAB sets tool precision, recall and F1 to 0 when both tool sets are empty, where Eq. 6 sets them to 1. RMSE and tool metrics are otherwise defined as above.

B Full Results and Analysis

B.1 Baseline Details

All baselines use the Direct protocol's system prompt, tool catalogue, channel constraints and decision-object schema. The protocols below specify how each produces its decision.

Prompting Baselines. Direct generates a decision in one model call. Chain of thought adds a reasoning field ahead of the decision in the same call. Self-consistency draws five samples at temperature 0.7, takes a majority vote on the intervention side, and returns the highest-scoring sample on the selected side. ReAct interleaves thought, action and observation for at most three steps before a final answer, with tool calls sandboxed so that an observation records that the call was accepted and returns no live data. Reflexion drafts, critiques the draft against the task constraints, and revises, for at most two rounds, stopping early when the critique reports no problems. Reflexion relies on self-critique and makes no external verifier call. Inner Thoughts forms System 1 and System 2 candidate thoughts, rates each on the eight heuristics of Liu et al. (2025), and produces a response only when the best rating clears their threshold of 3.59, with its turn-taking and retrieval-decay stages omitted in the single-turn setting. Proactive Agent proposes the single most useful action, then judges as the user whether it would be accepted on need, timing and burden, and acts when the acceptance probability exceeds one half. For POISE's single-context input, the same model performs this judgment in place of the original trained reward model of Lu et al. (2025). Native reasoning at high effort uses the model's own reasoning mode at its highest setting. One sample therefore costs one call under Direct and chain of thought, two or three under Proactive Agent, three under Inner Thoughts, up to four under ReAct, five under self-consistency and up to five under Reflexion.

Floor Policies. Three policies ignore the input. *Always silent* returns a proactivity score of one, *always act* a score of five, and *random* intervenes with probability one half, drawing a score from $\{4, 5\}$ when it acts and from $\{1, 2\}$ otherwise. All three return empty tool sets and delivery plans. Their tool and channel scores follow the empty-set metric convention, providing input-independent reference values for each metric.

B.2 Prompted Models and Training-Free Methods

Metrics and abbreviations follow Appendix A.1. Tables 12 and 13 give the complete prompted-model results on POISE-Lite and the full split. Tables 14–16 and 17–19 give the corresponding training-free method results in three metric blocks per split. Within each model or setting block the best value is in bold and the second best underlined. A double dagger marks native reasoning at high effort run without JSON mode. Tables 20 and 21 report the three floor policies on CAB and CAB-Lite.

B.3 Reinforcement Learning

Reward Components. For a rollout o with proactivity score σ, tool set T, channel set C and named decisive variable v, and a state $\mathbf{s}$ with reference score $\sigma^\star$, tools $T^\star$ and channels $C^\star$, the six

Table 12: All metrics of the prompted models on POISE-Lite (Table 2).

(a) Decision and tools

Model	Size	POISE↑	Acc-P↑	MD↓	FD↓	RMSE↓	Tool-P↑	Tool-R↑	Tool-F1↑	Unp.↓
Qwen3-0.6B	0.6B	17.6	50.0	**0.0**	100.0	1.582	1.9	1.0	1.3	0.0
Qwen3-1.7B	1.7B	21.3	50.4	**0.0**	99.2	**1.576**	8.3	6.6	6.6	0.0
Qwen3-4B	4B	23.4	50.2	0.4	99.2	2.040	19.3	11.8	13.6	0.0
Qwen3-4B-Inst.-2507	4B	24.7	50.2	**0.0**	99.6	2.161	19.2	20.2	18.0	0.0
Qwen3-8B	8B	23.1	50.4	0.4	98.8	2.125	19.0	11.2	13.1	0.0
Qwen2.5-1.5B-Inst.	1.5B	34.9	49.0	72.5	**29.6**	2.692	**38.2**	**37.3**	**37.6**	69.0
Qwen2.5-3B-Inst.	3B	**39.0**	54.7	49.0	41.7	2.382	34.4	33.1	33.5	50.0
Qwen2.5-7B-Inst.	7B	33.1	**58.5**	10.9	72.1	1.763	25.3	16.8	18.7	0.0
Llama-3.2-3B-Inst.	3B	20.3	50.0	**0.0**	100.0	2.082	8.7	6.1	6.6	0.0
Llama-3.1-8B-Inst.	8B	24.2	50.4	2.4	96.8	2.087	18.8	14.3	14.9	3.0
Gemma-3-4B-it	4B	21.9	50.0	0.4	99.6	2.080	15.3	10.6	11.5	0.0
GLM-4-9B-0414	9B	24.6	49.8	6.9	93.5	2.007	15.1	13.6	12.9	6.0
Floor policies										
Always silent	–	*37.5*	*50.0*	*100.0*	*0.0*	*2.199*	*50.0*	*50.0*	*50.0*	*0.0*
Always act	–	*37.5*	*50.0*	*0.0*	*100.0*	*2.858*	*50.0*	*50.0*	*50.0*	*0.0*
Random	–	*42.8*	*50.0*	*47.8*	*52.2*	*2.264*	*50.0*	*50.0*	*50.0*	*0.0*

(b) Pairs, channels, safety and timing

Model	Size	Pair-C.↑	Chan.-P↑	Chan.-R↑	Chan.-F1↑	MSVR↓	Order.↓	Deadl.↓
Qwen3-0.6B	0.6B	0.0	26.5	15.8	18.9	16.4	0.0	2.881
Qwen3-1.7B	1.7B	0.8	37.8	23.3	27.3	**0.2**	0.0	3.977
Qwen3-4B	4B	0.8	31.3	31.3	28.9	1.0	0.0	**2.502**
Qwen3-4B-Inst.-2507	4B	0.4	34.3	30.4	30.1	**0.2**	0.0	2.744
Qwen3-8B	8B	1.2	28.7	30.7	27.6	2.6	0.0	5.113
Qwen2.5-1.5B-Inst.	1.5B	10.5	42.6	**44.3**	42.7	1.0	0.0	3.948
Qwen2.5-3B-Inst.	3B	**24.7**	**44.9**	42.6	**43.2**	3.2	0.0	11.343
Qwen2.5-7B-Inst.	7B	21.5	34.8	36.4	33.8	3.6	0.0	7.873
Llama-3.2-3B-Inst.	3B	0.0	30.6	23.1	24.6	1.2	1.5	4.313
Llama-3.1-8B-Inst.	8B	2.4	31.1	33.5	29.2	2.2	9.2	4.715
Gemma-3-4B-it	4B	0.4	28.4	27.8	25.8	5.3	0.0	5.964
GLM-4-9B-0414	9B	6.5	34.2	30.2	29.5	1.0	0.0	4.429
Floor policies								
Always silent	–	*0.0*	*50.0*	*50.0*	*50.0*	*0.0*	*0.0*	–
Always act	–	*0.0*	*50.0*	*50.0*	*50.0*	*0.0*	*0.0*	–
Random	–	*21.1*	*50.0*	*50.0*	*50.0*	*0.0*	*0.0*	–

(c) Evidence, escalation, compliance and cost

Model	Size	DVI↑	Conf.↓	Over-esc.↓	Compl.↓	Calls	Tok.	Sec.
Qwen3-0.6B	0.6B	7.5	0.221	39.3	0.0	1.0	6863	0.28
Qwen3-1.7B	1.7B	8.5	0.248	**3.4**	0.0	1.0	6918	0.46
Qwen3-4B	4B	11.2	0.273	40.9	0.0	1.0	6795	0.66
Qwen3-4B-Inst.-2507	4B	15.0	0.278	11.1	0.0	1.0	6809	0.45
Qwen3-8B	8B	**16.5**	0.278	84.4	0.0	1.0	6813	0.94
Qwen2.5-1.5B-Inst.	1.5B	10.6	0.417	14.6	0.0	2.6	22884	3.75
Qwen2.5-3B-Inst.	3B	7.3	0.311	6.9	0.8	2.3	19134	3.67
Qwen2.5-7B-Inst.	7B	15.8	0.205	51.8	0.0	1.0	6997	0.87
Llama-3.2-3B-Inst.	3B	9.1	0.250	14.4	2.0	1.0	6751	0.32
Llama-3.1-8B-Inst.	8B	10.8	0.173	53.9	0.2	1.2	8758	5.19
Gemma-3-4B-it	4B	8.9	0.294	44.1	0.0	1.0	7978	0.82
GLM-4-9B-0414	9B	12.0	0.257	16.2	4.7	1.2	8380	5.89
Floor policies								
Always silent	–	–	–	*0.0*	*0.0*	*0.0*	*0*	*0.00*
Always act	–	–	–	*0.0*	*0.0*	*0.0*	*0*	*0.00*
Random	–	–	–	*0.0*	*0.0*	*0.0*	*0*	*0.00*

components are

$$r_{\text{dec}} = \tfrac{1}{2}\max\big(0, 1 - |\sigma - \sigma^\star|/2\big) + \tfrac{1}{2}\mathbb{1}\big[(\sigma \geq 3) = (\sigma^\star \geq 3)\big], \qquad r_{\text{tool}} = \mathrm{F}_1(T, T^\star), \tag{24}$$

$$r_{\text{evid}} = \mathbb{1}\big[v \in V^\star(\mathbf{s})\big], \qquad r_{\text{chan}} = \mathrm{F}_1(C, C^\star), \tag{25}$$

with F_1 as in Eq. 6. The safety component r_{safe} is one when every proposed action reaches a worn device through a channel in $\mathcal{A}(\mathbf{s}) \setminus \mathcal{F}(\mathbf{s})$ with consent respected, and it is defined only when o proposes an action, so an empty response cannot collect it. The timing component r_{time} is one when the plan's tool chain completes within $\tau(\mathbf{s})$ under the per-layer latency model of Appendix A.1. The weights are $(w_{\text{dec}}, w_{\text{tool}}, w_{\text{evid}}, w_{\text{chan}}, w_{\text{safe}}, w_{\text{time}}) = (0.40, 0.20, 0.15, 0.10, 0.10, 0.05)$. Each rollout is scored against the conditions of its own state by the evaluation scorer.

Table 13: All metrics of the prompted models on the full POISE test split (Table 2).

(a) Decision and tools

Model	Size	POISE↑	Acc-P↑	MD↓	FD↓	RMSE↓	Tool-P↑	Tool-R↑	Tool-F1↑	Unp.↓
Qwen3-0.6B	0.6B	18.1	50.0	0.1	99.9	**1.557**	3.1	1.7	2.2	0.0
Qwen3-1.7B	1.7B	21.5	50.2	0.1	99.7	**1.557**	9.3	6.8	7.1	0.0
Qwen3-4B	4B	23.4	50.3	0.1	99.4	2.007	19.9	11.6	13.6	0.0
Qwen3-4B-Inst.-2507	4B	25.4	50.3	0.1	99.3	2.170	20.9	21.3	19.4	0.0
Qwen3-8B	8B	24.3	49.3	6.0	95.4	2.100	14.9	9.2	10.5	5.0
Qwen2.5-1.5B-Inst.	1.5B	19.5	49.7	2.0	98.7	1.600	4.0	2.1	2.6	1.6
Qwen2.5-3B-Inst.	3B	19.1	50.9	0.9	97.3	1.587	6.6	2.8	3.9	0.2
Qwen2.5-7B-Inst.	7B	**30.3**	**54.7**	12.7	**78.0**	1.789	24.9	15.8	17.9	0.0
Llama-3.2-3B-Inst.	3B	20.8	50.1	**0.0**	99.8	2.029	8.2	5.2	5.9	0.0
Llama-3.1-8B-Inst.	8B	24.5	50.6	2.8	96.1	2.063	19.1	13.4	14.6	3.0
Gemma-3-4B-it	4B	22.6	50.0	0.5	99.4	2.037	16.0	10.5	11.7	0.0
GLM-4-9B-0414	9B	29.3	50.4	4.6	94.6	1.971	**33.7**	**30.2**	**30.5**	2.0
Floor policies										
Always silent	–	*37.5*	*50.0*	*100.0*	*0.0*	*2.230*	*50.0*	*50.0*	*50.0*	*0.0*
Always act	–	*37.5*	*50.0*	*0.0*	*100.0*	*2.805*	*50.0*	*50.0*	*50.0*	*0.0*
Random	–	*44.5*	*51.7*	*49.6*	*47.1*	*2.166*	*50.0*	*50.0*	*50.0*	*0.0*

(b) Pairs, channels, safety and timing

Model	Size	Pair-C.↑	Chan.-P↑	Chan.-R↑	Chan.-F1↑	MSVR↓	Order.↓	Deadl.↓
Qwen3-0.6B	0.6B	0.1	29.0	16.7	20.3	12.6	0.0	2.190
Qwen3-1.7B	1.7B	0.4	**39.7**	23.9	28.4	**0.2**	0.0	3.319
Qwen3-4B	4B	0.6	33.1	30.1	29.2	1.0	0.0	2.464
Qwen3-4B-Inst.-2507	4B	0.7	35.8	30.9	31.3	**0.2**	0.0	**2.188**
Qwen3-8B	8B	4.0	35.7	33.0	33.1	4.6	0.0	6.819
Qwen2.5-1.5B-Inst.	1.5B	1.2	26.2	27.6	24.6	3.9	0.4	9.346
Qwen2.5-3B-Inst.	3B	2.7	25.9	16.1	18.9	5.7	0.0	65.499
Qwen2.5-7B-Inst.	7B	**14.8**	34.4	**36.2**	**33.7**	1.9	0.0	11.526
Llama-3.2-3B-Inst.	3B	0.2	34.5	24.3	26.8	0.6	2.1	76.449
Llama-3.1-8B-Inst.	8B	3.3	31.4	33.0	29.4	2.0	13.4	7.036
Gemma-3-4B-it	4B	0.6	31.4	29.7	28.0	4.5	0.0	4.791
GLM-4-9B-0414	9B	4.7	35.4	30.5	31.5	5.5	0.1	3.633
Floor policies								
Always silent	–	*0.0*	*50.0*	*50.0*	*50.0*	*0.0*	*0.0*	–
Always act	–	*0.0*	*50.0*	*50.0*	*50.0*	*0.0*	*0.0*	–
Random	–	*26.2*	*50.0*	*50.0*	*50.0*	*0.0*	*0.0*	–

(c) Evidence, escalation, compliance and cost

Model	Size	DVI↑	Conf.↓	Over-esc.↓	Compl.↓	Calls	Tok.	Sec.
Qwen3-0.6B	0.6B	6.1	0.210	37.5	0.0	1.0	6796	0.17
Qwen3-1.7B	1.7B	9.5	0.239	5.3	0.0	1.0	6846	0.23
Qwen3-4B	4B	8.7	0.260	42.4	0.0	1.0	6782	0.31
Qwen3-4B-Inst.-2507	4B	10.6	0.268	12.4	0.0	1.0	6850	0.38
Qwen3-8B	8B	6.5	0.274	29.6	5.1	1.4	11107	12.24
Qwen2.5-1.5B-Inst.	1.5B	10.0	0.292	76.7	0.0	1.1	7553	0.62
Qwen2.5-3B-Inst.	3B	10.8	0.280	38.7	0.1	1.0	6903	0.39
Qwen2.5-7B-Inst.	7B	**13.1**	0.206	57.1	0.0	1.0	6806	0.35
Llama-3.2-3B-Inst.	3B	8.9	0.249	12.3	2.9	1.0	6756	0.24
Llama-3.1-8B-Inst.	8B	9.4	0.171	52.1	0.0	1.2	8580	3.98
Gemma-3-4B-it	4B	8.5	0.275	44.9	0.1	1.0	8019	0.38
GLM-4-9B-0414	9B	3.2	0.273	**4.6**	1.6	1.4	10188	4.72
Floor policies								
Always silent	–	–	–	*0.0*	*0.0*	*0.0*	*0*	*0.00*
Always act	–	–	–	*0.0*	*0.0*	*0.0*	*0*	*0.00*
Random	–	–	–	*0.0*	*0.0*	*0.0*	*0*	*0.00*

Optimization Objective. With $o_{i,t}$ the t-th token of rollout i and token-level importance ratio $\rho_{i,t}(\theta) = \pi_\theta(o_{i,t} \mid x, o_{i,<t}) / \pi_{\theta_{\text{old}}}(o_{i,t} \mid x, o_{i,<t})$, the reinforcement learning stage maximizes

$$\mathcal{J}(\theta) = \mathbb{E}\Big[\frac{1}{\sum_{i \in G} |o_i|} \sum_{i \in G} \sum_{t=1}^{|o_i|} \Big(\min\big(\rho_{i,t} A_i,\ \text{clip}(\rho_{i,t}, 1-\varepsilon, 1+\varepsilon) A_i\big) - \beta\, \hat{\mathbb{D}}_{\text{KL}}^{(i,t)} \Big)\Big], \tag{26}$$

where $A_i = (R_i - \text{mean}_G\, R)/(\text{std}_G\, R + \epsilon)$ is shared by all tokens of rollout i, $\hat{\mathbb{D}}_{\text{KL}}^{(i,t)}$ is the per-token estimate of the KL divergence to π_{ref} applied as a loss term with $\beta = 0.01$, and the loss is averaged over all response tokens in the batch.

Training Configuration. The supervised stage updates all parameters for five epochs at a learning rate of 10^{-5} under a cosine schedule with a warmup ratio of 0.1, an effective batch of eight

Table 14: All metrics of the training-free methods on POISE-Lite (Table 3). Part a: decision and tools. Parts b and c are Tables 15 and 16.

Model	Method	POISE↑	Acc-P↑	MD↓	FD↓	RMSE↓	Tool-P↑	Tool-R↑	Tool-F1↑	Unp.↓
Qwen3-8B	Direct	23.1	50.4	0.4	98.8	2.125	19.0	11.2	13.1	0.0
	CoT	24.5	51.8	**0.0**	96.4	2.246	21.9	12.9	15.1	0.0
	SC	23.1	50.4	0.4	98.8	2.263	18.3	10.8	12.6	0.0
	ReAct	23.3	50.6	0.8	98.0	2.193	20.6	11.3	13.6	0.0
	Reflexion	23.3	50.6	0.4	98.4	2.121	19.2	11.1	13.1	0.0
	Inner Th.	22.6	50.0	**0.0**	100.0	2.822	19.0	11.3	13.3	0.0
	Proact.	**30.0**	**54.5**	4.1	**87.0**	2.153	**32.7**	**20.2**	**23.1**	1.0
	Reason.-high‡	24.2	51.6	1.6	95.1	**2.080**	23.7	14.2	16.5	1.0
Qwen3-4B	Direct	23.4	50.2	**0.4**	99.2	2.040	19.3	11.8	13.6	0.0
	CoT	26.1	51.8	3.6	92.7	2.075	21.6	13.6	15.5	0.0
	SC	22.6	49.8	**0.4**	100.0	2.088	19.1	11.6	13.4	0.0
	ReAct	22.7	51.0	1.2	96.8	2.119	17.0	8.7	10.7	0.0
	Reflexion	23.3	50.2	**0.4**	99.2	2.037	19.0	11.7	13.4	0.0
	Inner Th.	25.2	51.0	1.6	96.4	2.523	21.7	13.8	15.6	0.0
	Proact.	**38.6**	**52.0**	79.8	**16.2**	2.150	**46.5**	**44.4**	**44.8**	0.0
	Reason.-high‡	27.1	**52.0**	4.5	91.5	**1.936**	24.6	14.4	16.8	1.0
Qwen2.5-7B-Inst.	Direct	**33.1**	**58.5**	10.9	**72.1**	**1.763**	25.3	**16.8**	**18.7**	0.0
	CoT	30.2	55.7	8.1	80.6	1.881	27.2	15.2	17.8	0.0
	SC	29.8	56.1	12.6	75.3	2.004	21.3	13.9	15.6	0.0
	ReAct	31.3	56.7	14.6	**72.1**	1.771	28.9	11.6	14.8	0.0
	Reflexion	28.8	55.7	4.9	83.8	1.779	21.1	12.3	14.3	2.0
	Inner Th.	23.0	50.4	**0.0**	99.2	2.615	21.2	11.2	13.5	0.0
	Proact.	23.3	50.4	2.0	97.2	1.812	**29.4**	10.8	14.6	0.0
	Reason.-high	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Qwen2.5-1.5B-Inst.	Direct	34.9	49.0	72.5	29.6	2.692	38.2	37.3	37.6	69.0
	CoT	36.6	49.2	89.5	12.2	2.892	44.5	44.1	44.2	88.1
	SC	38.3	51.2	55.5	42.1	2.498	35.5	34.5	34.7	45.8
	ReAct	**39.0**	**51.8**	82.2	14.2	2.799	43.7	43.6	43.7	82.0
	Reflexion	37.3	50.0	92.3	**7.7**	2.922	**46.7**	**46.7**	**46.7**	91.9
	Inner Th.	32.6	48.2	51.0	52.6	2.912	27.2	25.1	25.5	48.0
	Proact.	18.8	50.2	**2.0**	97.6	**1.620**	13.2	5.0	6.6	**1.6**
	Reason.-high	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Qwen3-1.7B	Direct	21.3	50.4	**0.0**	99.2	**1.576**	8.3	6.6	6.6	**0.0**
	CoT	28.1	49.4	22.3	79.0	1.966	18.1	17.4	16.9	20.5
	SC	21.2	50.2	**0.0**	99.6	1.591	9.7	6.9	7.3	**0.0**
	ReAct	29.1	49.8	25.1	75.3	2.014	**22.4**	17.0	18.1	24.3
	Reflexion	24.4	51.0	5.7	92.3	1.776	12.6	10.7	10.7	6.7
	Inner Th.	**29.6**	51.0	24.7	**73.3**	2.308	22.0	**18.4**	**19.2**	**0.0**
	Proact.	24.2	**51.4**	5.7	91.5	1.641	15.4	11.1	12.1	0.6
	Reason.-high‡	23.1	51.2	0.4	97.2	1.671	20.6	11.6	13.8	**0.0**

sequences, bfloat16 weights and an eight-bit AdamW optimizer, with prompts truncated at 7,680 tokens. The reinforcement learning stage runs GRPO for two epochs over the training split, roughly 640 optimization steps, drawing 16 prompts per step and five rollouts per prompt at temperature 1.0 and top-p 1.0, with an actor batch of eight, a learning rate of 10^{-5} under AdamW with weight decay 10^{-2}, and a low-variance estimator of the KL term of Eq. 4 at $\beta = 0.01$ against the supervised policy and a clipping range of $\varepsilon = 0.2$. Prompts are truncated at 7,168 tokens and responses at 768. Validation decodes at temperature 0.6 and top-p 0.95 with a single sample. The reward-component ablation of Table 5 reruns this configuration once per retained subset of terms, changing only which components Eq. 3 sums over.

B.4 Analysis

Prompt Composition. Table 23 varies the prompt of DeepSeek-V3.2, sampled at temperature 0.2. Default is zero-shot with the fused context, the persona and a decision-evidence field in the output. Each other condition changes one element. On the input side, w/o Personas drops the persona, Raw Context replaces the fused context with the wording of an on-device perception stack, and + Signals appends the structured signal snapshot. On the output side, w/o Evidence drops the decision-evidence field, and Base Fields omits the deadline, confidence, decisive-variable and abstention fields, which leaves DVI and the confidence gap undefined. Finally, 3-shot adds three training demonstrations.

Table 15: Table 14, continued. Part b: pairs, channels, safety and timing.

Model	Method	Pair-C.↑	Chan.-P↑	Chan.-R↑	Chan.-F1↑	MSVR↓	Order.↓	Deadl.↓
Qwen3-8B	Direct	1.2	28.7	30.7	27.6	2.6	0.0	5.113
	CoT	3.6	28.1	30.7	27.6	3.0	0.0	9.015
	SC	1.2	29.1	**31.3**	28.0	2.6	0.0	**4.764**
	ReAct	2.0	30.0	27.8	27.1	3.2	0.0	10.097
	Reflexion	1.6	29.0	31.0	27.9	2.6	0.0	5.728
	Inner Th.	0.0	27.9	29.8	27.1	3.0	0.0	4.825
	Proact.	**11.3**	**35.6**	30.5	**31.1**	**0.0**	0.0	11.409
	Reason.-high‡	4.5	24.8	26.5	24.1	1.4	0.0	19.532
Qwen3-4B	Direct	0.8	31.3	31.3	28.9	1.0	0.0	2.502
	CoT	5.7	33.6	33.8	31.5	2.0	0.0	9.003
	SC	0.0	29.7	29.3	27.1	1.6	0.0	2.876
	ReAct	2.8	29.7	27.1	26.5	3.2	0.0	6.362
	Reflexion	0.8	31.4	31.1	28.8	1.0	0.0	2.606
	Inner Th.	3.2	34.0	32.5	31.0	0.8	0.0	2.912
	Proact.	**9.7**	**48.6**	**47.9**	**47.7**	**0.0**	0.0	**1.344**
	Reason.-high‡	7.7	36.5	31.8	31.8	0.6	0.0	6.665
Qwen2.5-7B-Inst.	Direct	**21.5**	34.8	**36.4**	**33.8**	3.6	0.0	7.873
	CoT	15.8	33.4	33.6	31.7	1.4	2.3	9.969
	SC	17.8	30.5	32.8	29.8	3.4	1.5	8.077
	ReAct	**21.5**	**38.4**	30.3	32.1	**0.4**	0.0	7.171
	Reflexion	15.0	30.9	33.1	30.1	4.1	0.0	8.143
	Inner Th.	0.8	28.4	30.4	27.4	2.4	0.0	**6.698**
	Proact.	2.0	33.4	23.9	26.1	0.6	1.5	11.401
	Reason.-high	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Qwen2.5-1.5B-Inst.	Direct	10.5	42.6	44.3	42.7	1.0	0.0	3.948
	CoT	6.5	46.4	47.2	46.5	0.6	0.0	0.956
	SC	**22.7**	44.5	45.3	44.5	4.1	0.0	1.312
	ReAct	13.0	47.7	48.0	47.6	1.8	0.0	16.831
	Reflexion	4.1	**48.0**	**49.4**	**48.4**	**0.0**	0.0	**0.677**
	Inner Th.	22.3	35.7	35.0	34.3	4.3	0.0	15.854
	Proact.	2.0	19.8	15.8	16.3	6.3	0.8	11.738
	Reason.-high	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Qwen3-1.7B	Direct	0.8	37.8	23.3	27.3	0.2	0.0	**3.977**
	CoT	12.6	**42.9**	**30.1**	**33.7**	**0.0**	0.0	6.300
	SC	0.4	36.8	23.2	27.0	0.8	0.0	4.838
	ReAct	16.6	38.4	29.1	31.7	1.0	0.0	5.926
	Reflexion	6.9	37.8	25.8	28.9	0.8	0.0	4.276
	Inner Th.	**17.0**	39.5	28.3	31.4	0.4	0.0	4.441
	Proact.	7.7	33.3	22.9	25.7	1.4	0.0	6.509
	Reason.-high‡	2.8	31.1	22.7	24.7	**0.0**	0.0	4.075

B.5 Per-Group Breakdowns

Tables 24 and 25 break the full-split leaderboard down by scenario and by latency tier. Breakdowns for every run, metric, level and suppression predicate are released with the code.

Table 16: Table 14, continued. Part c: evidence, escalation, compliance and cost.

Model	Method	DVI↑	Conf.↓	Over-esc.↓	Compl.↓	Calls	Tok.	Sec.
Qwen3-8B	Direct	16.5	0.278	84.4	0.0	1.0	6813	0.94
	CoT	16.3	0.275	83.8	0.0	1.0	6933	1.93
	SC	15.0	0.279	84.4	0.0	5.0	34071	6.00
	ReAct	15.7	0.273	65.2	1.2	3.7	25454	4.69
	Reflexion	**17.4**	0.278	84.6	0.0	2.3	15505	2.74
	Inner Th.	15.4	0.273	83.0	0.0	3.0	9204	3.36
	Proact.	15.6	0.304	**54.5**	0.0	2.9	7894	3.09
	Reason.-high‡	15.5	0.264	75.5	0.2	1.1	8802	17.34
Qwen3-4B	Direct	11.2	0.273	40.9	0.0	1.0	6795	0.66
	CoT	10.8	0.270	32.2	0.0	1.0	6889	0.64
	SC	11.4	0.271	42.5	0.0	5.0	34016	1.98
	ReAct	9.7	0.274	10.3	0.0	3.6	24391	2.03
	Reflexion	11.0	0.275	39.5	0.0	2.0	13722	0.84
	Inner Th.	11.3	0.272	35.8	0.0	3.0	8971	1.78
	Proact.	4.5	0.523	**7.9**	0.0	1.4	2106	0.35
	Reason.-high‡	**11.9**	0.265	22.7	0.0	1.1	8363	5.86
Qwen2.5-7B-Inst.	Direct	15.8	0.205	51.8	0.0	1.0	6997	0.87
	CoT	15.1	0.244	40.9	0.0	1.0	7308	1.50
	SC	15.7	0.214	48.6	0.0	5.0	34102	2.14
	ReAct	15.5	0.215	20.0	0.0	3.6	24785	1.26
	Reflexion	**17.0**	0.203	58.5	0.0	2.4	16161	1.26
	Inner Th.	12.4	0.214	54.7	0.0	3.0	9288	1.89
	Proact.	15.5	0.208	**6.5**	0.0	3.0	8198	1.33
	Reason.-high	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Qwen2.5-1.5B-Inst.	Direct	10.6	0.417	14.6	0.0	2.6	22884	3.75
	CoT	**12.5**	0.600	5.1	0.0	2.9	25771	4.42
	SC	7.3	0.298	**3.0**	2.8	9.5	79539	9.32
	ReAct	3.9	0.242	6.1	0.0	3.8	32139	4.46
	Reflexion	6.5	0.383	4.1	0.0	3.4	30093	3.49
	Inner Th.	8.4	0.317	35.6	0.2	3.3	12429	4.23
	Proact.	10.4	0.323	24.7	0.2	3.0	8690	1.15
	Reason.-high	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Qwen3-1.7B	Direct	8.5	0.248	**3.4**	0.0	1.0	6918	0.46
	CoT	8.4	0.280	4.9	0.0	1.6	13195	7.63
	SC	8.3	0.248	8.1	0.0	5.1	34312	1.08
	ReAct	11.0	0.265	8.7	0.6	2.4	18857	8.60
	Reflexion	9.1	0.248	6.9	0.0	5.0	35679	3.01
	Inner Th.	7.4	0.249	4.1	0.0	2.7	7556	0.72
	Proact.	**12.9**	0.269	17.0	0.2	2.9	8071	1.26
	Reason.-high‡	12.0	0.265	13.2	0.2	1.0	8270	2.03

Table 17: All metrics of the training-free methods on the full POISE test split (Table 3). Part a: decision and tools. Parts b and c are Tables 18 and 19.

Model	Method	POISE↑	Acc-P↑	MD↓	FD↓	RMSE↓	Tool-P↑	Tool-R↑	Tool-F1↑	Unp.↓
Qwen3-8B	Direct	24.3	49.3	6.0	95.4	2.100	14.9	9.2	10.5	5.0
	CoT	24.8	51.4	0.6	96.6	2.241	22.8	12.6	15.1	0.0
	SC	23.4	50.4	0.3	99.0	2.223	19.9	11.2	13.4	0.0
	ReAct	22.8	50.7	0.5	98.2	2.198	22.5	11.8	14.4	0.0
	Reflexion	23.5	50.2	0.4	99.3	2.091	20.8	11.7	14.0	0.0
	Inner Th.	23.1	50.0	**0.0**	100.0	2.770	21.2	11.6	14.0	0.0
	Proact.	**29.0**	**53.3**	5.0	**88.4**	2.170	**32.9**	**19.0**	**22.2**	0.0
	Reason.-high‡	24.1	51.3	0.6	96.9	**2.079**	25.3	13.9	16.8	0.0
Qwen3-4B	Direct	23.4	50.3	**0.1**	99.4	2.007	19.9	11.6	13.6	0.0
	CoT	27.0	52.6	2.6	92.3	2.025	21.5	13.0	15.1	0.0
	SC	23.4	50.3	**0.1**	99.4	2.082	19.7	11.5	13.5	0.0
	ReAct	23.1	50.7	1.2	97.4	2.128	17.6	9.1	11.2	1.0
	Reflexion	23.4	50.3	**0.1**	99.5	2.008	19.9	11.6	13.6	0.0
	Inner Th.	24.8	50.8	1.0	97.4	2.513	21.2	12.5	14.6	0.0
	Proact.	**39.8**	**53.5**	74.0	**18.9**	2.109	**46.4**	**43.3**	**44.0**	0.0
	Reason.-high‡	26.4	52.0	3.1	93.0	**1.911**	22.6	12.8	15.2	0.0
Qwen2.5-7B-Inst.	Direct	30.3	54.8	12.6	77.9	1.789	24.9	**15.7**	**17.9**	0.0
	CoT	28.5	53.6	8.4	84.5	1.841	25.3	13.6	16.3	0.0
	SC	28.9	54.1	12.4	79.4	1.980	22.3	13.9	15.9	0.0
	ReAct	**30.5**	**55.4**	14.7	**74.7**	**1.777**	28.7	11.9	15.2	0.0
	Reflexion	26.2	52.6	5.1	89.6	1.797	20.7	11.0	13.3	0.0
	Inner Th.	22.8	50.2	**0.3**	99.4	2.583	20.2	10.1	12.5	0.0
	Proact.	23.6	50.5	2.0	97.0	1.807	**31.3**	11.3	15.3	0.0
	Reason.-high	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Qwen2.5-1.5B-Inst.	Direct	19.5	49.7	**2.0**	98.7	**1.600**	4.0	2.1	2.6	1.6
	CoT	37.5	49.9	96.9	3.3	3.011	48.7	48.7	48.7	96.3
	SC	**39.9**	**50.0**	68.5	31.4	2.657	40.5	40.2	40.3	58.4
	ReAct	38.5	**50.0**	92.9	7.2	2.962	48.3	48.1	48.2	90.7
	Reflexion	37.6	**50.0**	98.5	**1.5**	3.027	**49.4**	**49.4**	**49.4**	98.5
	Inner Th.	33.7	49.4	50.2	51.0	2.916	27.0	25.5	25.9	48.8
	Proact.	24.5	49.9	2.9	97.3	1.675	16.9	8.3	9.8	**1.3**
	Reason.-high	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Qwen3-1.7B	Direct	21.5	50.2	0.1	99.7	**1.557**	9.3	6.8	7.1	0.1
	CoT	29.4	50.7	21.6	77.1	1.980	18.2	17.5	17.1	21.1
	SC	21.5	50.1	**0.0**	99.9	1.578	9.8	6.9	7.4	**0.0**
	ReAct	29.3	49.9	27.4	**72.8**	2.054	21.7	17.2	18.2	26.6
	Reflexion	23.7	50.2	6.0	93.7	1.748	12.5	10.1	10.3	6.2
	Inner Th.	**31.1**	**51.5**	23.8	73.3	2.256	**22.3**	**18.8**	**19.6**	**0.0**
	Proact.	23.7	51.1	5.9	91.9	1.612	14.3	9.8	10.8	0.4
	Reason.-high‡	23.0	50.5	0.8	98.3	1.678	21.2	11.5	13.8	0.1

Table 18: Table 17, continued. Part b: pairs, channels, safety and timing.

Model	Method	Pair-C.↑	Chan.-P↑	Chan.-R↑	Chan.-F1↑	MSVR↓	Order.↓	Deadl.↓
Qwen3-8B	Direct	4.0	35.7	**33.0**	**33.1**	4.6	0.0	6.819
	CoT	3.4	30.4	31.4	29.2	2.2	0.0	7.691
	SC	0.9	30.4	31.3	29.0	2.5	0.0	5.171
	ReAct	1.6	26.1	25.8	24.3	2.5	0.0	7.491
	Reflexion	0.6	30.9	31.6	29.3	1.9	0.0	5.131
	Inner Th.	0.0	29.9	30.3	28.4	2.3	0.0	**4.855**
	Proact.	**9.0**	**36.6**	30.8	31.4	**0.4**	0.0	8.813
	Reason.-high‡	3.0	26.4	27.4	25.4	1.3	0.0	8.047
Qwen3-4B	Direct	0.6	33.1	30.1	29.2	1.0	0.0	2.464
	CoT	7.1	36.4	34.4	33.2	1.8	0.0	4.879
	SC	0.6	33.7	29.8	29.3	1.3	0.0	2.745
	ReAct	2.6	31.3	27.7	27.7	3.5	0.0	4.078
	Reflexion	0.5	33.2	30.0	29.2	1.0	0.0	3.399
	Inner Th.	2.5	36.4	31.2	31.3	0.5	0.0	5.348
	Proact.	**13.8**	**49.7**	**47.7**	**48.0**	**0.0**	0.0	**1.705**
	Reason.-high‡	6.2	38.1	31.5	32.3	0.9	0.0	7.982
Qwen2.5-7B-Inst.	Direct	15.0	34.4	**36.2**	**33.6**	1.9	**0.0**	11.506
	CoT	12.3	33.1	33.6	31.6	**0.3**	2.4	10.640
	SC	14.2	33.1	33.5	31.6	2.1	2.2	22.726
	ReAct	**18.1**	**39.1**	31.6	33.3	**0.3**	**0.0**	**8.760**
	Reflexion	8.9	30.8	32.8	29.9	1.9	0.1	12.236
	Inner Th.	0.6	28.8	30.7	27.8	0.9	2.2	11.121
	Proact.	2.3	33.3	24.3	26.4	0.4	0.8	11.699
	Reason.-high	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Qwen2.5-1.5B-Inst.	Direct	1.2	26.2	27.6	24.6	3.9	0.4	9.346
	CoT	2.0	49.4	49.7	49.4	**0.1**	0.0	**1.614**
	SC	21.9	47.6	47.5	47.4	1.3	0.0	–
	ReAct	6.4	49.5	49.5	49.4	**0.1**	0.0	–
	Reflexion	1.2	**49.7**	**50.0**	**49.8**	**0.1**	0.0	8.006
	Inner Th.	**22.9**	38.4	37.1	36.7	2.0	0.1	10.886
	Proact.	2.3	38.6	35.0	35.8	1.8	0.3	8.262
	Reason.-high	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Qwen3-1.7B	Direct	0.4	39.7	23.9	28.4	0.2	0.0	3.319
	CoT	15.3	**43.4**	30.8	**34.4**	0.2	0.0	3.307
	SC	0.1	39.5	24.1	28.5	0.3	0.0	4.191
	ReAct	17.5	38.6	29.0	31.8	1.0	0.0	4.583
	Reflexion	5.7	37.7	24.9	28.4	0.5	0.3	**3.105**
	Inner Th.	**19.0**	42.7	**31.0**	**34.4**	**0.1**	0.0	4.038
	Proact.	7.1	33.9	22.3	25.7	0.8	0.0	3.930
	Reason.-high‡	1.7	34.0	23.2	26.1	0.4	0.0	3.695

Table 19: Table 17, continued. Part c: evidence, escalation, compliance and cost.

Model	Method	DVI↑	Conf.↓	Over-esc.↓	Compl.↓	Calls	Tok.	Sec.
Qwen3-8B	Direct	6.5	0.274	**29.6**	5.1	1.4	11107	12.24
	CoT	12.6	0.265	79.4	0.1	1.0	7305	1.62
	SC	12.4	0.268	81.6	0.0	5.0	34134	3.26
	ReAct	**13.3**	0.277	74.0	0.0	3.8	26292	2.11
	Reflexion	12.9	0.268	81.1	0.0	2.2	15342	1.11
	Inner Th.	12.6	0.263	80.5	0.0	3.0	9232	2.12
	Proact.	11.9	0.280	53.1	0.1	2.9	7882	1.44
	Reason.-high‡	12.0	0.255	77.0	0.0	1.0	7862	3.68
Qwen3-4B	Direct	8.7	0.260	42.4	0.0	1.0	6782	0.31
	CoT	9.1	0.260	35.3	0.0	1.0	6885	0.38
	SC	8.9	0.262	44.6	0.0	5.0	33915	1.24
	ReAct	**9.5**	0.266	**10.4**	0.0	3.6	24747	1.07
	Reflexion	8.7	0.261	42.8	0.0	2.0	13737	0.43
	Inner Th.	7.9	0.262	40.4	0.0	3.0	9055	1.21
	Proact.	6.0	0.486	10.5	0.0	1.5	2449	0.31
	Reason.-high‡	**9.5**	0.260	26.7	0.0	1.0	7642	1.45
Qwen2.5-7B-Inst.	Direct	13.0	0.208	57.1	0.0	1.0	6790	0.36
	CoT	11.9	0.238	43.0	0.0	1.0	6997	0.51
	SC	12.8	0.216	45.1	0.0	5.0	33897	1.24
	ReAct	12.0	0.222	23.1	0.0	3.6	24372	0.66
	Reflexion	13.1	0.208	63.4	0.0	2.3	15588	0.53
	Inner Th.	11.7	0.216	57.4	0.0	3.0	8999	1.11
	Proact.	**13.3**	0.209	**9.7**	0.0	3.0	8104	0.79
	Reason.-high	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Qwen2.5-1.5B-Inst.	Direct	**10.0**	0.292	76.7	0.0	1.1	7553	0.62
	CoT	6.9	0.066	1.2	0.0	3.0	26461	3.93
	SC	3.0	0.290	1.0	1.3	9.7	80449	7.62
	ReAct	8.8	0.157	0.7	0.0	3.9	33274	4.26
	Reflexion	2.7	0.043	0.8	0.0	3.1	27755	3.40
	Inner Th.	9.8	0.291	35.7	0.0	3.3	12585	3.89
	Proact.	8.2	0.319	**0.0**	2.0	3.0	7982	0.28
	Reason.-high	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Qwen3-1.7B	Direct	9.5	0.239	5.3	0.0	1.0	6846	0.23
	CoT	9.9	0.272	5.5	0.0	1.6	13418	7.79
	SC	9.6	0.243	7.9	0.0	5.0	34249	0.99
	ReAct	11.0	0.260	9.3	0.7	2.4	18747	8.82
	Reflexion	9.7	0.240	7.7	0.0	5.0	35395	2.67
	Inner Th.	8.3	0.243	**3.9**	0.0	2.7	7613	0.66
	Proact.	**11.4**	0.258	19.5	0.1	2.9	7986	1.00
	Reason.-high‡	9.9	0.261	16.2	0.1	1.0	8067	1.40

Table 20: Floor policies on CAB.

Policy	Index↑	MD↓	FD↓	Score↑	RMSE↓	Tool-P↑	Tool-R↑	Tool-F1↑
Always silent	*47.0*	*53.0*	*0.0*	*43.5*	*2.610*	*46.8*	*46.8*	*46.8*
Always act	*53.0*	*0.0*	*47.0*	*31.5*	*2.759*	*46.8*	*46.8*	*46.8*
Random	*50.5*	*25.0*	*24.5*	*23.5*	*2.416*	*46.8*	*46.8*	*46.8*

Table 21: Floor policies on CAB-Lite.

Policy	Index↑	MD↓	FD↓	Score↑	RMSE↓	Tool-P↑	Tool-R↑	Tool-F1↑
Always silent	*55.8*	*44.2*	*0.0*	*53.3*	*2.388*	*55.8*	*55.8*	*55.8*
Always act	*44.2*	*0.0*	*55.8*	*26.7*	*3.006*	*55.8*	*55.8*	*55.8*
Random	*45.8*	*25.0*	*29.2*	*24.2*	*2.497*	*55.8*	*55.8*	*55.8*

Table 22: All metrics of SFT and SFT with the PACT reward on Qwen2.5-1.5B-Instruct (Table 4).

Benchmark	Metric	SFT	+GRPO	Δ
POISE	POISE↑	83.1	89.0	+5.9
	Acc-P↑	88.7	93.3	+4.6
	MD↓	6.7	3.5	−3.2
	FD↓	16.0	9.9	−6.0
	RMSE↓	1.089	0.864	−0.225
	Tool-P↑	83.5	89.6	+6.1
	Tool-R↑	83.0	87.4	+4.4
	Tool-F1↑	81.4	86.8	+5.4
	Unp.↓	0.1	0.0	−0.1
	Pair-C.↑	78.7	86.9	+8.2
	Chan.-P↑	84.7	90.5	+5.9
	Chan.-R↑	84.4	89.0	+4.6
	Chan.-F1↑	83.7	89.0	+5.3
	MSVR↓	0.3	0.1	−0.2
	Order.↓	0.0	0.0	+0.0
	Action.↑	99.7	99.8	+0.1
	Deadl.↓	0.041	0.043	+0.002
	DVI↑	64.8	66.5	+1.7
	Abst.-P↑	97.4	97.5	+0.0
	Abst.-R↑	95.0	96.7	+1.7
	Conf.↓	0.011	0.011	+0.000
	Over-esc.↓	0.0	0.0	+0.0
	Compl.↓	0.0	0.0	+0.0
CAB	Index↑	91.5	92.3	+0.8
	MD↓	3.5	5.3	+1.8
	FD↓	5.0	2.5	−2.5
	Score↑	65.5	65.8	+0.3
	RMSE↓	1.091	1.063	−0.028
	Tool-P↑	75.8	76.8	+1.1
	Tool-R↑	75.1	75.8	+0.7
	Tool-F1↑	73.6	74.7	+1.0

Table 23: Prompt composition for DeepSeek-V3.2 (DeepSeek-AI et al., 2025) on the full POISE test split. Conf.: confidence gap. Tok.: tokens per sample, in thousands. **Bold**: best per column.

Condition	POISE↑	Acc-P↑	MD↓	FD↓	Pair-C.↑	Tool-F1↑	Chan.-F1↑	DVI↑	MSVR↓	Conf.↓	RMSE↓	Tok.
Default	35.0	55.6	18.6	70.3	22.8	22.3	39.2	17.0	0.2	33.9	2.279	7.1
w/o Personas	33.5	54.5	14.4	76.8	19.1	22.5	38.1	15.7	0.2	31.8	2.283	7.0
Raw Context	34.2	54.4	17.6	73.7	18.1	23.3	41.3	16.2	0.1	23.0	2.044	6.8
+ Signals	35.9	55.5	18.2	70.7	22.6	25.3	40.3	17.2	**0.0**	40.3	2.351	7.3
w/o Evidence	35.8	50.7	91.2	**7.4**	7.8	36.5	**48.0**	**27.7**	**0.0**	**18.7**	3.797	6.6
Base Fields	32.7	52.6	**12.6**	82.2	14.7	25.0	38.5	–	0.1	–	2.321	6.7
3-shot	**42.9**	**61.0**	14.8	63.3	**27.5**	**37.2**	46.0	23.0	0.1	23.1	**1.793**	8.4

Table 24: Leaderboard models on the full POISE test split by scenario, grouped by latency tier. Top: POISE score. Bottom: pair consistency. **Bold**: best model per scenario. Column names follow Table 7.

	Reflex		Interactive							Deferred		
Model	Rear	Cross	Elder	Focus	Heat	Talk	Venue	Gate	Lang	Airport	Child	Meet
POISE score												
Qwen3-0.6B	20.4	13.6	23.2	16.3	22.6	16.2	18.4	14.0	20.3	15.6	17.0	13.7
Qwen3-1.7B	20.2	24.1	28.0	17.2	22.0	24.3	23.3	17.6	23.6	16.2	20.6	14.4
Qwen3-4B	25.4	28.5	25.5	17.3	24.7	29.6	23.2	21.4	26.1	18.0	20.9	16.5
Qwen3-4B-Inst.-2507	27.1	27.6	28.5	19.3	24.8	29.3	25.5	24.3	29.1	19.5	**28.2**	18.8
Qwen3-8B	23.6	27.6	25.1	17.7	24.0	26.6	24.8	21.8	27.9	24.2	25.0	18.8
Qwen2.5-7B-Inst.	**33.8**	**32.7**	29.1	**31.3**	26.5	**33.9**	27.1	29.3	**40.8**	**29.9**	23.1	**37.0**
Llama-3.2-3B-Inst.	19.2	22.7	25.5	18.0	21.2	25.6	22.1	13.8	17.9	16.8	21.5	15.5
Llama-3.1-8B-Inst.	26.5	26.9	22.9	20.3	29.4	26.7	25.3	22.1	30.5	19.5	23.2	19.2
Gemma-3-4B-it	22.7	26.7	25.1	17.0	23.9	27.5	25.5	17.8	16.9	19.3	21.9	15.9
GLM-4-9B-0414	30.1	31.7	**29.7**	29.5	**30.5**	29.3	**28.6**	**29.5**	31.6	27.8	26.6	27.8
Always silent	37.5	37.5	37.5	37.3	37.5	37.5	37.5	37.5	37.5	37.5	37.5	37.5
Always act	37.5	37.5	37.5	37.4	37.5	37.5	37.5	37.5	37.5	37.5	37.5	37.5
Random	45.5	46.9	45.4	48.4	43.1	42.9	45.4	43.5	43.8	42.4	44.4	41.7
Pair consistency												
Qwen3-0.6B	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Qwen3-1.7B	0.7	0.7	0.0	0.0	0.0	0.0	0.5	1.4	2.8	0.0	0.0	0.0
Qwen3-4B	0.0	2.9	0.0	0.0	1.0	0.6	0.5	0.0	0.0	0.5	0.0	1.2
Qwen3-4B-Inst.-2507	2.0	0.7	0.0	0.0	1.0	0.6	0.0	2.8	0.0	0.0	0.0	3.5
Qwen3-8B	1.3	8.1	2.0	0.0	3.1	1.3	1.1	5.6	8.3	8.1	**9.2**	2.3
Qwen2.5-7B-Inst.	**24.0**	**11.8**	**7.1**	**25.0**	7.1	**12.6**	**8.7**	**26.8**	**36.1**	**19.5**	0.0	**34.5**
Llama-3.2-3B-Inst.	0.7	0.0	0.0	0.0	0.5	0.0	0.0	0.0	2.8	0.0	0.0	0.0
Llama-3.1-8B-Inst.	1.3	5.9	0.5	1.0	**14.7**	0.0	1.1	1.4	4.2	0.0	4.9	2.3
Gemma-3-4B-it	0.7	0.0	1.0	0.0	0.0	0.0	1.1	7.0	0.0	0.0	0.0	0.0
GLM-4-9B-0414	5.8	**11.8**	1.5	6.7	5.1	2.5	5.4	8.5	4.2	3.2	1.8	4.6
Always silent	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Always act	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Random	27.3	30.9	29.3	33.7	24.9	22.6	28.1	26.8	25.0	20.5	25.0	20.7

Table 25: Leaderboard models on the full POISE test split, broken down by latency tier. **Bold**: best model per column.

	Reflex				Interactive				Deferred			
Model	Acc-P↑	Pair-C.↑	Chan.-F1↑	POISE↑	Acc-P↑	Pair-C.↑	Chan.-F1↑	POISE↑	Acc-P↑	Pair-C.↑	Chan.-F1↑	POISE↑
Qwen3-0.6B	50.0	0.0	13.8	17.2	50.0	0.1	26.1	19.4	50.0	0.0	11.5	15.7
Qwen3-1.7B	50.2	0.7	31.5	22.1	50.2	0.4	33.6	23.1	50.0	0.0	14.5	17.5
Qwen3-4B	50.5	1.4	35.7	26.9	50.2	0.4	31.6	24.5	50.2	0.5	19.3	18.8
Qwen3-4B-Inst.-2507	50.5	1.4	34.4	27.3	50.3	0.5	33.2	26.1	50.3	0.7	24.9	22.7
Qwen3-8B	49.8	4.5	**35.8**	25.5	49.2	2.4	**34.6**	24.3	49.2	7.3	28.2	23.4
Qwen2.5-7B-Inst.	**50.9**	**18.3**	34.8	**33.3**	**56.0**	**13.7**	34.3	**30.1**	**54.4**	**15.1**	**31.5**	**28.8**
Llama-3.2-3B-Inst.	50.2	0.3	24.6	20.8	50.2	0.3	31.0	21.8	50.0	0.0	18.5	18.3
Llama-3.1-8B-Inst.	50.7	3.5	34.1	26.7	50.8	3.8	32.4	25.5	50.0	2.3	19.7	20.8
Gemma-3-4B-it	50.0	0.3	30.4	24.6	50.1	0.9	29.4	23.3	50.0	0.0	23.4	19.6
GLM-4-9B-0414	50.7	8.6	32.0	30.9	50.7	4.4	33.6	29.7	49.5	3.0	26.4	27.4
Always silent	50.0	0.0	50.0	37.5	50.0	0.0	50.0	37.5	50.0	0.0	50.0	37.5
Always act	50.0	0.0	50.0	37.5	50.0	0.0	50.0	37.5	50.0	0.0	50.0	37.5
Random	55.5	29.0	50.0	46.1	51.3	27.1	50.0	44.6	49.8	22.3	50.0	43.0

C CASE STUDIES

Figures 10–15 group fourteen baseline cases from all twelve scenarios by failure type. Each panel shows the situation, decisive variable, reference and model outputs, and scored verdict. Figure 16 adds three observed SFT+PACT responses, complementing the four response forms in Fig. 6.

(a) Acting on both halves of a pair

Qwen2.5-7B-Instruct, Inner Thoughts, both halves • talk pacing • INTERACTIVE

INTERVENING HALF *"Speaker identification attributes the captured voice to the presenter rather than to anyone in the audience."*

SILENT HALF *"The speech under measurement is not the presenter's own … another person in the room is the one talking."*

Both halves share a speech rate of 204 words per minute, 33.9 percent above the speaker's baseline, sustained for 64 seconds.

DECISIVE VARIABLE `speech.is_user_speaking`, the only variable that differs.

REFERENCE	MODEL
Intervening half (score 4): a watch haptic and "Slow down" on the glasses. Silent half (score 1): stay silent.	Intervening half (score 5): a haptic and "Rushing" on the glasses. Silent half (score 5): a haptic and "The speaker is going too fast" spoken from the watch.

VERDICT Channel F1 1.00 (intervening half) | Pair consistency 0

The silent half also breaks the no-speech constraint of a live talk and escalates to an emergency call.

(b) Reversed pair decisions

Qwen2.5-7B-Instruct, Self-consistency, both halves • venue norms • INTERACTIVE

INTERVENING HALF *"Within fifty metres of the cinema now: the entrance is 2.0 m off and the user is walking at 1.27 m/s."*

SILENT HALF *"At 1.27 m/s the cinema door is 138.0 m ahead, well beyond fifty metres."*

In both halves the user, a self-declared smoker, has an appointment at the non-smoking cinema in 35 minutes.

DECISIVE VARIABLE `poi.dist_m`, the only variable that differs.

REFERENCE	MODEL
Intervening half (score 4): a soft haptic, "Whole site is non-smoking. Nearest area is 80 m east." through the earbuds, and "No smoke" on the glasses. Silent half (score 1): stay silent.	Intervening half (score 2): stays silent. Silent half (score 3): a haptic, a warning on the phone, and a message to a contact.

VERDICT Both halves wrong | Pair consistency 0

The intervention decision changes in the direction opposite to the distance predicate.

Figure 10: **Counterfactual pair failures.** (a) Intervention in both halves. (b) Decisions reverse in the direction opposite to the distance predicate. Pair consistency is zero in both.

(a) Correct decision, inaccessible channel

Qwen2.5-7B-Instruct, Direct, intervening half • airport flow • DEFERRED

"Wrist IMU output is walking with luggage and the wrist is not free to be raised. … There is a booking in the user's own name, on time, boarding in 49.0 minutes, with the gate 297.0 m away."

DECISIVE VARIABLE `flight.minutes_to_boarding`, named correctly by the model.

REFERENCE ACT 5

Spoken directions through the earbuds, "Straight on 120 metres, then right to gate C21", a gate marker on the glasses, and a single soft haptic on the watch.

MODEL ACT 3

A double haptic and a screen card on the watch, then earbuds. The wrist is occupied with luggage and cannot be raised, so the watch screen is the one surface the wearer cannot look at.

VERDICT Decision correct | Channel F1 0.67 | Tool F1 0.00

Correct intervention does not imply correct tool or channel selection.

(b) Safety violation from a corrupted plan

Qwen2.5-7B-Instruct, Direct, intervening half • elder routine • INTERACTIVE

"11.0 minutes have passed without movement, compared with the 1.0 minutes of stillness that is ordinary at this hour. The expected departure came and went 12.0 minutes ago. At 2.4 sigma, that lateness sits outside two standard deviations of the learned time. … This morning walk has been learned across 22 days."

DECISIVE VARIABLE `routine.adherence_rate`: the person normally keeps to this routine, so a departure is meaningful.

REFERENCE ACT 4

One very soft haptic on the watch, intensity 0.3.

MODEL ACT 4

Decides to act, and emits a device field reading `watchwatchwatch{watchwatch` with the plan nested inside it.

VERDICT Decision correct | Channel F1 0.00 | Safety `unavailable_device`

The decision object parses, yet its plan reaches no device.

(c) Channel-order violation

Qwen2.5-7B-Instruct, CoT, intervening half • rear approach • REFLEX

"With ANC engaged at 0.93, the earbuds are removing the very sound that would otherwise announce this. … From the rear, the mic array tracks something approaching at 9.1 m/s … Yaw of 3.0 deg means the wearer has not looked back."

DECISIVE VARIABLE `earbuds.head_yaw_deg`: the rider has not looked back.

REFERENCE ACT 5

The earbuds ramp to transparency with a sustained haptic on the watch, and "Vehicle behind." follows through the earbuds 200 ms later.

MODEL ACT 4

"A vehicle is approaching from behind. Please look back." through the earbuds at 0 ms, with noise cancellation lowered at the same instant, plus an escalation that calls emergency services.

VERDICT Decision correct | Channel F1 0.80 | Ordering `earbuds.anc->earbuds.tts` | Over-escalation

Speech begins before the plan reduces noise cancellation, violating the required channel order.

Figure 11: **Correct decisions with delivery errors.** (a) A cue uses an inaccessible display. (b) The plan reaches no device. (c) Channels execute in the wrong order.

(a) Correct decision, incorrect evidence

Qwen2.5-7B-Instruct, Direct, silent half • airport flow • DEFERRED

"There are 145.0 minutes left on the boarding clock … The gate is 10.0 m away, and the position estimate sits in the gate area on level 1. Positioning uncertainty reads 47.0 m with a one-floor error in the level estimate, too coarse and too displaced to route from."

DECISIVE VARIABLES `indoor.accuracy_m`, `indoor.floor_error`: the fix is not good enough to route from.

REFERENCE SILENT 1

Stay silent.

MODEL SILENT 2

Stays silent, and names `watch.imu.wrist_free` as the reason.

VERDICT Decision correct | Decisive variable missed

The decision is correct, but the evidence omits the variable that determines the label.

(b) Chain of thought repairs the Direct decision

Qwen3-4B, CoT versus Direct, silent half • elder routine • INTERACTIVE

"On record from the person is an explicit note that they are not going out today. The walk itself had begun and then stalled partway … Across 21 days the device has learned 'morning walk' as a fixed piece of the morning, kept 0.86 of the time."

DECISIVE VARIABLE `routine.cancelled_by_user`: the person has already said they are not going out.

REFERENCE SILENT 1

Stay silent.

DIRECT AND COT

Direct acts. CoT gives score 2 and stays silent, reasoning that "the user is in a familiar location and the door is locked from inside", and names `indoor.zone` as decisive.

VERDICT Decision repaired | Decisive variable missed

The revised decision is correct, but its stated evidence differs from the reference condition.

(c) Correct decision and evidence

Qwen2.5-7B-Instruct, Direct, intervening half • cross-language assist • INTERACTIVE

"The user is in Japan and the language on the far side of the conversation is German, outside the Chinese-native, English-fluent pair their profile records. The geometry (0.9 m, face to face) and the timing of the turns across one speaker both say the same thing: the user is who is being spoken to … A translation app is not open … "

DECISIVE VARIABLE `speech.language`, named correctly by the model.

REFERENCE ACT 4

"Ice?" on the glasses, and "They asked if you want ice." softly through the earbuds.

MODEL ACT 4

Starts German-to-English translation and speaks through the earbuds, without the glasses cue.

VERDICT Decision correct | Decisive variable named | Channel F1 0.67 | Tool F1 0.33

The decision and decisive variable match the reference, while tool and channel selection remain incomplete.

Figure 12: **Decision evidence and delivery.** (a) Correct decision with missing decisive evidence. (b) Revised decision with evidence outside the reference condition. (c) Correct decision and evidence with incomplete tool and channel selection.

(a) Majority-vote false intervention

Qwen2.5-7B-Instruct, Self-consistency, silent half • road crossing • REFLEX

"The walk indicator at the crossing reads red. … A lit display carries a messaging app, and the scroll stream is still moving … Cadence has dropped clean away from baseline … and the user is now stationary at the kerb."

DECISIVE VARIABLES `crossing.user_stopped`, `watch.imu.cadence_hz`: the user has already stopped at the kerb.

REFERENCE SILENT 1

Stay silent.

MODEL ACT 5

All five samples vote to act: three haptics on the watch, "Watch the crossing signal." through earbuds that are not worn, a red icon on the watch, and an emergency call.

VERDICT False intervention | Safety `unavailable_device` | Over-escalation | Decisive variable missed

Majority voting retains the incorrect intervention.

(b) Named suppression, false intervention

Qwen2.5-7B-Instruct, Direct, silent half • focus shield • INTERACTIVE

"This class of item has been queued 4 separate times in the last seven days, with only 8% of such items acknowledged over 30 days. … By the geofence the user stands inside site_high_voltage, where the risk level is high … "

DECISIVE VARIABLES `ctx.same_alert_count_7d`, `ctx.user_ack_rate_30d`: the user has stopped engaging with this class of item.

REFERENCE SILENT 2

Stay silent.

MODEL ACT 4

Names `ctx.same_alert_count_7d` as decisive, then acts anyway, deferring safety notifications and raising the safety alert threshold to 4.

VERDICT False intervention | Decisive variable named

The model names the suppressing signal but still intervenes.

Figure 13: **Suppression-condition failures.** (a) Self-consistency votes to intervene after the pedestrian stops. (b) The model identifies the suppression variable but does not condition its decision on it.

(a) Resolved heat condition

Qwen2.5-7B-Instruct, Direct, silent half • heat stress • INTERACTIVE

"Over the last 9.0 minutes the heart-rate reading has settled at 130.0 bpm . . . Outdoor air is 38.9 C and 60.0 percent humid, an apparent temperature of 49.7 C, a level at which heat illness can develop. The user, however, has already come in out of it. GPS reports indoors and shade coverage of 0.94 . . . "

DECISIVE VARIABLES `phone.gps.indoors`, `phone.gps.shade_ratio`: indoor status suppresses additional assistance.

REFERENCE SILENT 1

Stay silent.

MODEL ACT 4

A strong haptic and a screen card on the watch, a phone call to an emergency contact, and an escalation that calls emergency services, with shelter distance named as decisive.

VERDICT False intervention | Over-escalation | Decisive variable missed

The model acts on elevated physiological readings without applying the suppressing indoor-status condition.

(b) Acting on an untrustworthy transcript

Qwen2.5-7B-Instruct, Direct, silent half • pa announcement • INTERACTIVE

"A voice on the terminal PA delivered a gate change. The earbud microphone array detected it. The number carried in the message is the user's flight. Transcription confidence sits at 0.31. At 1.2 dB of signal-to-noise under a 3.4 s decay, the individual words came apart and the recovered text cannot be trusted."

DECISIVE VARIABLES `pa.asr_conf`, `pa.snr_db`, `pa.rt60_s`: transcription confidence is insufficient for a gate-change alert.

REFERENCE SILENT 1

Stay silent.

MODEL ACT 4

"Your flight has a gate change. New gate is 368.0 meters away." on the earbuds, watch and glasses, and the same message sent to family.

VERDICT False intervention | Decisive variable missed

The model treats announcement detection as decisive despite insufficient transcription reliability.

Figure 14: **Evidence that should have blocked the intervention.** (a) the user has already resolved the condition, (b) an untrustworthy transcript is relayed as fact.

(a) Escalation despite low battery

Qwen2.5-7B-Instruct, Direct, silent half • child safety • DEFERRED

> *"Connectivity to the child's watch has been absent for 29.0 minutes. The last reported position was the metro entrance. It was down to 6.0 percent when it last transmitted, so power exhaustion covers the silence. . . ."*
>
> DECISIVE VARIABLE `child.last_battery_pct`: a nearly empty battery explains the silence.

REFERENCE SILENT 2

Stay silent.

MODEL ACT 4

A haptic on the guardian's watch, a message with location to the guardian, and an escalation that calls emergency services once the outage passes 30 minutes, with outage duration named as decisive.

VERDICT False intervention | Over-escalation | Decisive variable missed

The model mentions the 6 percent battery but does not use it as a suppression condition.

(b) Missed meeting-mode intervention

Qwen2.5-7B-Instruct, Direct, intervening half • meeting prep • DEFERRED

> *"The calendar entry is still standing, no cancellation recorded against it, and 2 attendees are listed on it. With 11.0 minutes left before the start, the meeting is near enough for readying to be a live option. . . . a beacon match from the meeting room has landed."*
>
> DECISIVE VARIABLE `calendar.cancelled`: the meeting still stands.

REFERENCE ACT 5

Switch the phone to meeting mode and hold incoming notifications until the meeting ends.

MODEL SILENT 2

Stays silent, with the time since the last dismissal named as decisive.

VERDICT Missed intervention | Decisive variable missed

The model gives an unrelated interaction record more weight than the meeting signals.

Figure 15: **Over-escalation and omission.** (a) A low battery leads to an unwarranted emergency call. (b) Meeting signals are present, but an unrelated interaction record determines the decision.

(a) False work interruption

Qwen3-1.7B, SFT+PACT vs SFT • focus shield • INTERACTIVE

The worker is inside a high-voltage site, but the wrist has been still for 12 seconds. Heart rate is 71 bpm against a 68 bpm baseline, RMSSD is 40 ms against 41 ms, and the queued group message has general priority.

REFERENCE Stay silent (score 1). The resting physiology does not warrant a new work-interruption policy.

PACT SILENT 1

No tools or output plan.

SFT ACT 5

A strong watch haptic and an earbuds message call the general-priority group message urgent.

VERDICT Correct silence | Tool F1 1.00 | Channel F1 1.00

PACT does not convert the non-actionable pause into an alert.

(b) Single-channel walking cue

Qwen3-1.7B, SFT+PACT vs Outcome • elder routine • INTERACTIVE

The morning walk started but stopped. The user did not cancel it, and 18 minutes of stillness exceed the usual 4.3 minutes. The departure is 13 minutes later than expected, or 2.32 standard deviations from the learned routine.

REFERENCE Act (score 3) with one very soft watch haptic.

PACT ACT 3

Queries the routine profile and sends one very soft watch haptic at 0 ms.

OUTCOME ACT 3

Sends a watch haptic, then adds a watch screen message and watch speech.

VERDICT Decision correct | Tool F1 1.00 | Channel F1 1.00

The PACT plan matches the reference tool and channel without extra output modalities.

(c) Speech-pacing cue during a talk

Qwen3-1.7B, SFT+PACT vs SFT • talk pacing • INTERACTIVE

The presenter speaks at 248 words per minute, 38.3 percent above their baseline, for 62 seconds. Speaker identification attributes the voice to the user, and the previous pacing hint was 146 seconds ago.

REFERENCE Act (score 4) with a watch haptic and “Slow down” on the glasses.

PACT ACT 3

Sends a soft watch haptic and a glasses text cue reading “Rate” at 0 ms.

SFT ACT 3

Sends a watch haptic without a glasses cue.

VERDICT Decision correct | Tool F1 0.67 | Channel F1 0.50

PACT adds a visual cue, although its wording is less specific than the reference.

Figure 16: Additional SFT+PACT cases: silence, a watch cue, and a two-device pacing cue.