

# From Latent Empathy to Structured Training Signals for Emotional Support Dialogue

Linran Xu

Tsinghua University

Shenzhen International Graduate School

xulr24@mails.tsinghua.edu.cn

Xiaomei Nie

Tsinghua University

Shenzhen International Graduate School

nie.xiaomei@sz.tsinghua.edu.cn

## Abstract

Emotional support conversation (ESC) requires responses that recognize users’ feelings and situational meaning while balancing relational support with practical orientation. Yet it remains unclear whether improvements from synthetic post-training reflect structured empathy signals or merely stronger rewritten responses. We introduce structured empathic supervision, representing each turn with *Feeling/Meaning* understanding targets and *Subjective/Interpersonal/Objective* (S/I/O) response plans. On a coverage-oriented diagnostic suite of 219 turns from 20 ES-Conv dialogues, Full-SIO-SFT significantly outperforms a decoding- and budget-matched Rewrite-SFT control on aggregate support quality, perceived empathy, counseling quality, and cognitive depth. Controlled perspective analyses show target-aligned specialization: SI favors relational support, whereas IO favors informational support. Starting from Full-SIO-SFT, balanced DPO further improves the aggregate six-dimensional support objective, while directional objectives induce distinct support profiles. Blinded comparative ratings from two psychology-background annotators corroborate the structured-supervision gain and the directional profile shift. These results support structured supervision and preference weighting as complementary mechanisms for controllable ESC post-training.

## 1 Introduction

Emotional support conversation (ESC) requires more than producing responses that sound warm. A supportive system must recognize what a user feels, understand why the situation matters to them, and respond in a way that is appropriate to the evolving dialogue context. A response can therefore appear empathetic while still being ineffective if it misses the user’s underlying concern, offers premature advice, or becomes overly directive. ES-

Conv formalized this challenge as a strategy-aware multi-turn support task (Liu et al., 2021).

Recent work improves ESC through stronger language models, explicit reasoning structures, supporter-intention modeling, and feedback-based training (Kang et al., 2024; Zhang et al., 2024b, 2025b; Wang et al., 2024). However, two questions remain open. First, when synthetic supervision improves ESC, it is often unclear whether gains arise from the proposed structure or merely from higher-quality rewritten responses. Second, emotional support is inherently multi-objective: empathy, sensitivity, non-directiveness, practical usefulness, and informational content may not improve together.

We study whether a theory-grounded structure can provide more effective post-training signals for ESC. Drawing on Clark’s integrative account of empathy (Clark, 2007, 2010), we represent each turn using two understanding targets—*Feeling* and *Meaning*—and three response perspectives: *Subjective*, *Interpersonal*, and *Objective* (S/I/O). *Feeling* captures the user’s affective state; *Meaning* captures the significance of the situation; S/I/O describe complementary ways of expressing support. We use this decomposition as a computational supervision scaffold, not as a claim that a model recovers psychological ground truth.

Our pipeline tests three questions. For structured supervision, we compare **Full-SIO-SFT** against a **Rewrite-SFT** control generated from the same source instances using the same teacher model, explicit decoding configuration, response-length constraint, target format, and supervised-training budget. The matched conditions differ in their target-construction structure: Full-SIO supplies S/I/O perspective plans together with S/I/O-specific integration guidance, whereas Rewrite retains *Feeling*, *Meaning*, and strategy information without S/I/O-specific conditioning. We additionally evaluate a prompt-only baseline supplied with the same structured information at inference time. For behavioral

validity, we train single-view S/I/O models and generate controlled SI, SO, IO, and Full-SIO candidates. For preference optimization, we construct DPO data from six support dimensions and train balanced and directionally weighted variants.

Across 25 systems evaluated on 219 turns from 20 independent dialogues, we find that Full-SIO-SFT significantly outperforms Rewrite-SFT across DPO6D, perceived empathy, CpsyCounEvalMatrix, and Carkhuff cognitive depth. Prompting recovers part of the structured behavior, especially on Interpersonal and Objective adherence, but remains weaker in overall quality and Subjective adherence. Perspective ablations and controlled candidates show that S/I/O are behaviorally meaningful: SI favors relational dimensions, while IO strongly favors informational support. Finally, DPO-Balanced significantly improves the DPO6D aggregate beyond Full-SIO-SFT, while directional weighting induces interpretable within-family profile shifts with complementary differences on independently formulated quality outcomes.

Our contributions are threefold:

- We formulate structured empathic supervision for ESC by separating Feeling/Meaning understanding targets from S/I/O response perspectives.
- We provide controlled evidence that Full-SIO supervision improves beyond decoding- and budget-matched rewriting and inference-time structured prompting.
- We show that DPO-Balanced yields a significant gain on the DPO6D aggregate beyond structured SFT, while directional objectives induce statistically distinguishable relational-informational response profiles.

## 2 Related Work

**Emotional support dialogue.** EmpatheticDialogues studies emotionally grounded open-domain response generation (Rashkin et al., 2019), while ESConv formalizes strategy-aware multi-turn emotional support (Liu et al., 2021). Recent systems improve support through language-model adaptation, explicit reasoning, intention modeling, multifaceted feedback, and reinforcement learning (Kang et al., 2024; Zhang et al., 2024b, 2025b; Wang et al., 2024; Yuan et al., 2025b,a). Kardia-R1 studies persona-grounded emotional support, while

ReflectDiffu focuses on empathetic response generation. Because these settings differ from situation-centric ESConv, we report Kardia-R1 only as an external reference rather than a matched baseline (Appendix I). Our focus is whether a theory-grounded empathy decomposition provides a useful post-training signal beyond decoding- and budget-matched synthetic rewriting.

**Structured empathy and preference optimization.** Prior work shows that empathy-oriented supervision and grounding can improve supportive generation (Chen et al., 2023; Qian et al., 2023), and that theoretical empathy constructs affect downstream modeling choices (Lahnala et al., 2025). Our Feeling/Meaning and S/I/O formulation operationalizes such structure for supervised training, controlled composition, and parameter-space analysis. We further build on DPO (Rafailov et al., 2023) and recent preference-based ESC work (Zhang et al., 2025a) by separating aggregate DPO improvement from the calibration effect of different support objectives.

**Evaluation.** ESC evaluation is multi-dimensional: ESC-Eval and CPsyCoun emphasize strategy-sensitive and counseling-oriented quality (Zhao et al., 2024; Zhang et al., 2024a), while perceived-empathy work emphasizes whether users feel understood and supported (Concannon and Tomalin, 2024; Yin et al., 2024). We therefore use four primary outcome families and validate LLM-based evaluation with human annotations, following broader concerns about judge reliability (Calderon et al., 2025; Li et al., 2025; Kumar et al., 2026).

## 3 Method

**Structured empathic representation.** Given a dialogue context and current user utterance  $x$ , we construct:

$$z(x) = (f, m, p_S, p_I, p_O, \sigma), \quad (1)$$

where  $f$  and  $m$  denote the target *Feeling* and *Meaning*,  $p_S$ ,  $p_I$ , and  $p_O$  denote Subjective, Interpersonal, and Objective response plans, and  $\sigma$  denotes the planned ESConv support strategy. Feeling captures the user’s affective state, whereas Meaning captures the personal significance of the situation. S/I/O define complementary expression channels: affective resonance, relational alignment, and grounded or practical orientation.

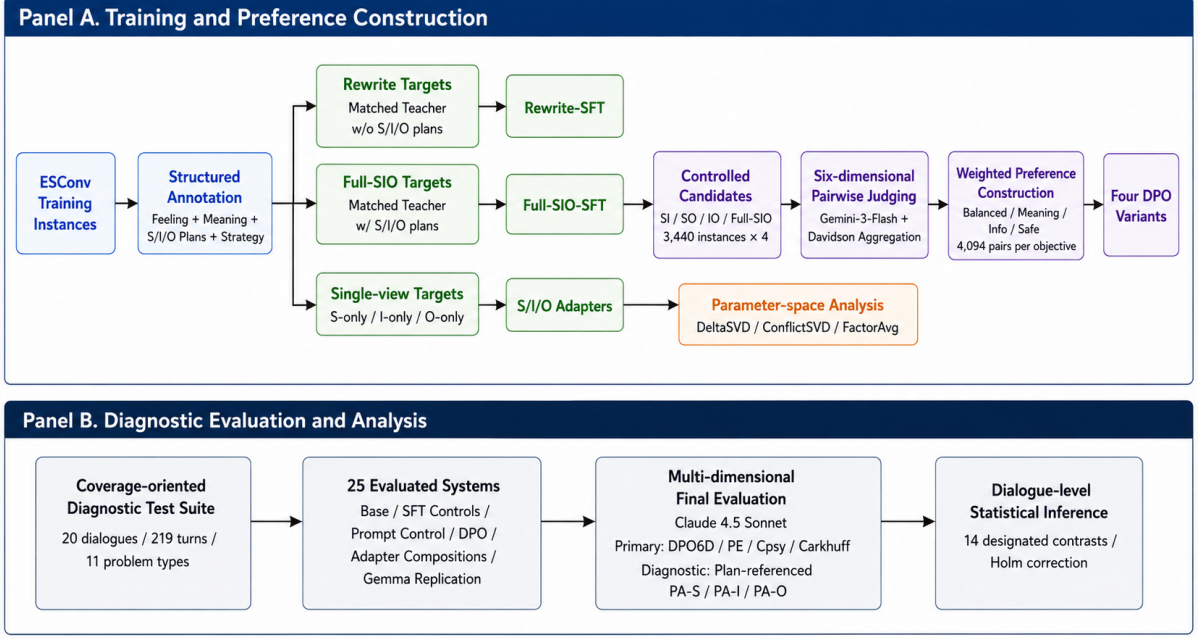


**Figure 1:** Overview of the controlled post-training and evaluation framework. Training instances are annotated with Feeling/Meaning targets and S/I/O perspective plans to construct decoding- and budget-matched Rewrite-SFT and Full-SIO-SFT supervision, single-view adapters, controlled perspective candidates, and four weighted DPO variants. A separate coverage-oriented diagnostic test suite is used for multi-dimensional evaluation and dialogue-level inference.

**Structured supervision and controls.** We synthesize 5,823 **Full-SIO** targets conditioned on all elements of  $z(x)$  and fine-tune the base model to obtain **Full-SIO-SFT**. As the principal controlled comparison, **Rewrite-SFT** uses the same source instances, teacher model, explicit decoding configuration, strategy information, output format, response-length constraint, and supervised-training budget, but omits S/I/O perspective-plan conditioning. Thus, Full-SIO-SFT versus Rewrite-SFT estimates the benefit of S/I/O-conditioned structured target construction over matched non-S/I/O rewriting within the synthetic-supervision pipeline. We additionally train **S-only**, **I-only**, and **O-only** models using the corresponding single-view plan, and evaluate a **Base + S/I/O Prompt** condition that supplies the full structure only at inference time.

**Controlled candidate composition.** Using Full-SIO-SFT, we generate four matched candidates for each of 3,440 training instances:

$$\mathcal{Y}(x) = \{y^{SI}, y^{SO}, y^{IO}, y^{\text{Full-SIO}}\}. \quad (2)$$

Candidates are generated by Full-SIO-SFT, not the teacher model, and share context, annotations, strategy, checkpoint, and decoding configuration; only their supplied perspective plans differ.

**Multi-objective preference construction.** We judge candidate pairs on six support dimensions: empathy, identification, effectiveness, sensitivity, non-directiveness, and informational support. For each of the 3,440 candidate sets, all six unordered candidate pairs are evaluated in both presentation orders on all six dimensions, yielding 247,680 dimension-level judgments. Preference judgments are produced by gemini-3-flash, which is separate from the Claude 4.5 Sonnet model used for final automatic evaluation.

For each support dimension  $d$ , we estimate a pooled Davidson tie parameter  $\hat{w}_d$  (Davidson, 1970) and, with it fixed, fit normalized candidate worths  $\tilde{s}_d(y | x)$  separately within each instance  $x$ ; worths are never compared across contexts. We then compute objective-specific utilities:

$$R_w(y | x) = \sum_{d \in \mathcal{D}} w_d \tilde{s}_d(y | x). \quad (3)$$

Given the resulting ranking, each instance contributes a Top1-versus-Bottom preference pair and, when Top1 is not Full-SIO, an additional Top1-versus-Full-SIO pair. We discard pairs whose chosen-to-rejected length ratio exceeds 1.3 and remove duplicates before matched-size sampling. For the four DPO variants reported in the main experiments, this procedure yields 4,094 preference

| Objective | Emp.  | Id.   | Eff.  | Sens. | NonDir. | Info. |
|-----------|-------|-------|-------|-------|---------|-------|
| Balanced  | .1667 | .1667 | .1667 | .1667 | .1667   | .1665 |
| Meaning   | .25   | .35   | .10   | .15   | .10     | .05   |
| Info      | .08   | .05   | .30   | .12   | .10     | .35   |
| Safe      | .15   | .10   | .05   | .35   | .30     | .05   |

**Table 1:** Preference-objective weights used for the four reported DPO variants. Balanced is uniform up to four-decimal residual rounding; Meaning prioritizes identification; Info prioritizes effectiveness and informational support; Safe prioritizes sensitivity and non-directiveness.

pairs per objective.

**DPO and parameter-space analysis.** The four reported DPO variants are initialized from the same Full-SIO-SFT checkpoint and trained under the same matched-size weighted-DPO configuration; they differ only in the objective weights used to construct their preference pairs. For DPO training, each chosen–rejected pair is associated with the same policy input; perspective-specific conditioning is used only to generate and rank candidate responses, not to provide different conditioning contexts within the DPO loss. We use DPO following [Rafailov et al. \(2023\)](#). Prompt templates, model-role assignments, preference-construction details, evaluation protocols, and complete reported results are provided in the appendix. Finally, we analyze LoRA ([Hu et al., 2022](#)) updates from S/I/O single-view adapters. We measure update similarity, test linear reconstruction of Full-SIO, and construct rank-preserving DeltaSVD compositions to examine whether perspective directionality persists in parameter space.

## 4 Experimental Setup

**Data and systems.** We use the English ESConv dataset ([Liu et al., 2021](#)). Structured SFT conditions each use 5,823 training instances, and controlled candidate construction uses 3,440 training instances. Our principal experiments use Qwen3-4B ([Yang et al., 2025](#)), covering Base, Original-SFT, Rewrite-SFT, Full-SIO-SFT, prompt-only, single-view, DPO, and adapter-composition systems. We additionally conduct a reduced replication from pretrained google/gemma-3-4b-pt ([Gemma Team, 2025](#)). Prompt templates, model-role assignments, preference-construction details, evaluation protocols, and complete reported results are provided in the appendix.

**Evaluation subset construction.** We evaluate on a coverage-oriented ESConv test subset of 20 dialogues and 219 turns covering all 11 problem

types. The subset is selected under fixed problem-type quotas from a 100-dialogue stratified superset constructed using dataset-intrinsic metadata and deterministic case-ID sampling only; no model generations or evaluation scores enter selection. This design prioritizes scenario coverage over natural-frequency estimation. All inference therefore treats dialogues, rather than turns, as independent units.

**Evaluation metrics.** We treat four rubric families as primary outcomes: DPO6D, Perceived Empathy (PE), CpsyCounEvalMatrix (Cpsy), and Carkhuff cognitive depth. DPO6D measures empathy, identification, effectiveness, sensitivity, non-directiveness, and informational support. PE measures perceived understanding and support; Cpsy measures counseling-oriented quality and safety; and FM-Depth evaluates recognition of the user’s Feeling and Meaning annotations. DPO6D is aligned with the six dimensions used for preference construction but is scored by a separate evaluator model; we therefore interpret DPO gains jointly with PE, Cpsy, Carkhuff, and blinded human ratings as complementary evidence of response quality. Perspective Adherence (PA) is a secondary diagnostic metric measuring realization of the S/I/O plans. All PA results reported in the paper use the corrected plan-referenced protocol.

**Blinded human comparative evaluation.** To obtain evaluator-independent evidence for the central system comparisons, we conduct a blinded comparative human evaluation on 50 turn-level instances stratified by dialogue stage and problem type. Two annotators with psychology backgrounds independently rate five anonymized responses per instance, generated by Rewrite-SFT, Full-SIO-SFT, DPO-Balanced, DPO-Safe, and DPO-Info. Response order is randomized within each instance, and annotators do not observe model identities. The human evaluation covers perceived empathy (PE-H), Carkhuff cognitive depth (Carkhuff-H), simplified counseling norms (Cpsy-H), objective-perspective adherence (PA-O-H), and a reduced two-dimension DPO6D-H diagnostic. Because the 50 sampled instances arise from 18 dialogues, inferential comparisons aggregate paired differences within dialogue before statistical testing.

**Structured-annotation audit.** Separately from the response-quality evaluation, two annotators with psychology backgrounds audited a stratified 60-turn sample of the Feeling/Meaning and S/I/O

| Model               | DPO6D        | PE           | Cpsy         | Carkhuff     | PA-S        | PA-I        | PA-O        |
|---------------------|--------------|--------------|--------------|--------------|-------------|-------------|-------------|
| Base                | 2.751        | 3.309        | 4.494        | 2.110        | 1.67        | 2.34        | 1.49        |
| Original-SFT        | 2.413        | 2.559        | 4.347        | 1.406        | —           | —           | —           |
| Rewrite-SFT         | 3.316        | 3.880        | 4.899        | 2.662        | 2.27        | 2.78        | 2.07        |
| Base + S/I/O Prompt | 3.241        | 4.256        | 5.176        | 3.069        | 2.55        | 3.11        | 2.28        |
| Full-SIO-SFT        | 3.585        | 4.320        | 5.740        | 3.196        | 2.82        | <b>3.20</b> | <b>2.31</b> |
| DPO-Balanced        | 3.829        | 4.681        | <b>6.061</b> | <b>3.648</b> | 2.53        | 2.81        | 1.68        |
| DPO-Meaning         | 3.736        | 4.586        | 5.915        | 3.626        | 2.59        | 2.65        | 1.64        |
| DPO-Info            | 3.564        | 4.239        | 5.588        | 3.119        | 2.05        | 2.47        | 2.08        |
| DPO-Safe            | <b>3.831</b> | <b>4.752</b> | 5.730        | 3.580        | <b>2.83</b> | 2.83        | 1.50        |

**Table 2:** Main results on the 20-dialogue diagnostic test suite. DPO6D, PE, Cpsy, and Carkhuff are primary outcomes; PA-S/I/O are secondary plan-referenced diagnostics.

annotations. Across seven fields, 91.7–100% of ratings were at least 4 on a five-point scale, and within-one-point agreement was 96.7–98.3%. Two models that were not used to create these annotations audited 150 additional items and showed the same high-acceptability pattern. Field-level results and error categories are reported in Appendix L.

**Evaluation and statistics.** Our primary automatic evaluation uses a fixed Claude 4.5 Sonnet judge configuration across systems. To test evaluator dependence, we also rerun the four primary metrics for the 14 principal systems with GPT-5.2 and Gemini-3.5-Flash, changing only the judge model while keeping the prompts, decoding, parsing, and dialogue-level testing protocol fixed. For the primary confirmatory analyses, we aggregate paired turn-level differences within each dialogue and use dialogue-cluster bootstrap confidence intervals, dialogue-level sign-flip permutation tests, and Holm–Bonferroni correction across 14 designated contrasts. The non-Claude robustness analyses use the same dialogue-level tests, with Holm correction applied separately to the four metrics for each judge. For the targeted human comparisons, the 50 rated instances are aggregated within their 18 originating dialogues, and Holm correction is applied separately across 10 pre-specified human contrasts. Pooled instance-level means are reported as descriptive summaries. Full evaluation and statistical details are provided in Appendices E, D, G, H, and M.

## 5 Results

Table 2 summarizes the main model progression, while Table 3 reports the selected dialogue-level confirmatory contrasts used to support the central claims across RQ1–RQ3.

| Contrast        | Metric   | $\Delta$ | $d_z$  | $p_{\text{Holm}}$ |
|-----------------|----------|----------|--------|-------------------|
| SIO–Rewrite     | DPO6D    | +0.268   | 0.869  | 0.0036            |
| SIO–Rewrite     | PE       | +0.442   | 1.147  | 0.0014            |
| SIO–Rewrite     | Cpsy     | +0.851   | 1.755  | 0.0013            |
| SIO–Rewrite     | Carkhuff | +0.491   | 1.099  | 0.0012            |
| Balanced–SIO    | DPO6D    | +0.244   | 0.823  | 0.0040            |
| Safe–SIO        | DPO6D    | +0.232   | 0.885  | 0.0040            |
| Safe–Info       | DPO6D    | +0.265   | 1.138  | 0.0009            |
| Safe–Info       | Carkhuff | +0.468   | 1.068  | 0.0018            |
| Safe–Info       | PA-S     | +0.755   | 2.167  | < 0.0001          |
| Safe–Info       | PA-O     | -0.571   | -1.983 | < 0.0001          |
| DeltaSI–DeltaIO | PA-S     | +0.196   | 0.564  | 0.0210            |
| DeltaSI–DeltaIO | PA-O     | -0.191   | -0.647 | 0.0210            |

**Table 3:** Selected dialogue-level confirmatory contrasts. Effect sizes are paired Cohen’s  $d_z$ ;  $p_{\text{Holm}}$  denotes Holm-corrected sign-flip permutation results.

### 5.1 RQ1: Does Structured Supervision Improve Beyond Rewriting?

Table 2 shows a clear controlled progression. Original-SFT underperforms Base on all four primary outcomes. Its targets are the short, strategy-labeled supporter turns in ESConv, whereas Rewrite-SFT and Full-SIO-SFT use matched synthetic targets from the same teacher. We therefore treat Full-SIO-SFT versus Rewrite-SFT as the central controlled comparison: the two conditions share data, teacher, decoding, response length, output format, and training budget, and differ in S/I/O conditioning. Full-SIO-SFT is higher on all four primary metrics, with particularly large gains on Cpsy ( $\Delta = +0.851$ ,  $d_z = 1.755$ ,  $p_{\text{Holm}} = 0.0013$ ) and Carkhuff depth ( $\Delta = +0.491$ ,  $d_z = 1.099$ ,  $p_{\text{Holm}} = 0.0012$ ). The same advantage remains significant on all four metrics under GPT-5.2 and Gemini-3.5-Flash ( $p_{\text{Holm}} \leq 0.003$ ; Appendix H).

Prompting with the same structured annotations also helps. Base + S/I/O Prompt is substantially stronger than Base and approaches Full-SIO-SFT on Carkhuff depth (3.069 vs. 3.196). However, it

| A. Single-view PA gain over Base |               |               |               | B. Candidate-level specialization |      |              |
|----------------------------------|---------------|---------------|---------------|-----------------------------------|------|--------------|
| Model                            | $\Delta$ PA-S | $\Delta$ PA-I | $\Delta$ PA-O | Dimension                         | View | Win Rate     |
| S-only                           | <b>+1.00</b>  | +0.62         | +0.50         | Empathy                           | SI   | 31.6%        |
| I-only                           | +0.87         | <b>+0.70</b>  | +0.54         | Identification                    | SI   | 31.6%        |
| O-only                           | +0.87         | +0.60         | <b>+0.73</b>  | Sensitivity                       | SI   | 30.6%        |
| Full-SIO                         | +1.15         | +0.86         | +0.82         | Non-directiveness                 | SI   | 29.9%        |
|                                  |               |               |               | Informational                     | IO   | <b>60.9%</b> |

**Table 4:** Behavioral evidence for perspective structure: single-view PA gains and candidate-level specialization.

remains below Full-SIO-SFT on every primary outcome. The plan-referenced PA results provide a more nuanced picture: prompting approaches Full-SIO-SFT on Interpersonal adherence (3.11 vs. 3.20) and Objective adherence (2.28 vs. 2.31), while retaining a clearer gap on Subjective adherence (2.55 vs. 2.82). Structured supervision therefore contributes beyond prompt-time access to the same information, with its clearest perspective-specific advantage appearing in the Subjective channel. Blinded human ratings independently corroborate this result: Full-SIO-SFT significantly exceeds Rewrite-SFT on PE-H, Carkhuff-H, and Cpsy-H after correction over ten targeted human contrasts (Appendix M).

## 5.2 RQ2: Are S/I/O Perspectives Behaviorally Meaningful?

**Target-aligned perspective learning.** Panel A of Table 4 shows that the S/I/O channels function as selective behavioral targets rather than interchangeable supervision variants. Among the single-view models, S-only-SFT obtains the largest Subjective gain over Base (+1.00), I-only-SFT obtains the largest Interpersonal gain (+0.70), and O-only-SFT obtains the largest Objective gain (+0.73). Rewrite-SFT improves all three channels to some extent, but does not exhibit the same target-aligned specialization pattern. Full-SIO-SFT further improves all three channels jointly, indicating that the single-view models exhibit target-aligned behavioral specialization while Full-SIO establishes broader perspective coverage.

**Composition is selective rather than monotonically additive.** Panel B shows a complementary candidate-level result. When SI, SO, IO, and Full-SIO responses are generated under matched context, strategy, and decoding conditions, SI is strongest on relational dimensions, including empathy, identification, sensitivity, and non-directiveness, whereas IO exhibits a pronounced advantage on informational support. Across can-

didate sets, SI is the most frequently top-ranked composition overall (33.0%), followed by Full-SIO (29.6%). Thus, joint Full-SIO supervision can be the strongest general training condition even when a more selective perspective composition is preferable for a particular response objective.

**Answer to RQ2.** The behavioral diagnostics provide consistent evidence that S/I/O conditioning induces differentiated response tendencies. Single-view supervision exhibits target-aligned PA patterns, while controlled candidate generation reveals a relational-informational contrast between SI and IO compositions. Because these behavioral analyses are used diagnostically rather than as confirmatory hypothesis tests, we interpret them as evidence of controllable differentiation rather than definitive proof of universally separable perspective mechanisms. Exploratory stage-wise trends are reported in Appendix J.

## 5.3 RQ3: Does DPO Improve and Calibrate Support?

**Preference optimization improves the aggregate support objective beyond structured SFT.** DPO-Balanced, which assigns approximately uniform weight to all six support dimensions, significantly improves DPO6D over Full-SIO-SFT ( $\Delta = +0.244$ ,  $d_z = 0.823$ ,  $p_{\text{Holm}} = 0.0040$ ). DPO-Safe provides a similarly reliable DPO6D gain over Full-SIO-SFT ( $\Delta = +0.232$ ,  $d_z = 0.885$ ,  $p_{\text{Holm}} = 0.0040$ ) and achieves the highest observed DPO6D and PE scores. Meanwhile, DPO-Balanced obtains the highest observed Cpsy and Carkhuff scores. Thus, preference optimization yields a statistically reliable improvement on the objective-aligned six-dimensional aggregate, while complementary automatic and human outcomes assess broader quality consequences beyond that aggregate.

**Directional weighting produces distinguishable response profiles.** The DPO variants do not con-

verge to the same behavioral solution. DPO-Safe preserves Subjective adherence at approximately the Full-SIO-SFT level (2.83 vs. 2.82) while reducing Objective adherence to approximately the Base level (1.50 vs. 1.49). Within the DPO family, DPO-Info exhibits the complementary shift: it obtains lower Subjective adherence than DPO-Safe (2.05 vs. 2.83) but higher Objective adherence (2.08 vs. 1.50). These profile differences are statistically reliable: DPO-Safe exceeds DPO-Info on PA-S ( $\Delta = +0.755$ ,  $d_z = 2.167$ ,  $p_{\text{Holm}} < 0.0001$ ), whereas DPO-Info exceeds DPO-Safe on PA-O ( $\Delta = -0.571$  for Safe minus Info,  $d_z = -1.983$ ,  $p_{\text{Holm}} < 0.0001$ ).

This directional redistribution is not a Pareto improvement over Full-SIO-SFT. In particular, DPO-Info does not exceed Full-SIO-SFT on PA-O, and DPO-Safe significantly exceeds DPO-Info on aggregate DPO6D ( $\Delta = +0.265$ ,  $d_z = 1.138$ ,  $p_{\text{Holm}} = 0.0009$ ) and Carkhuff cognitive depth ( $\Delta = +0.468$ ,  $d_z = 1.068$ ,  $p_{\text{Holm}} = 0.0018$ ). We therefore interpret directional objectives as controlling the realized balance among support qualities within the DPO family, rather than guaranteeing monotonic improvement of every targeted perspective over the structured-SFT initialization.

**Answer to RQ3.** Preference optimization serves two roles in ESC. First, balanced DPO yields a significant improvement on the automatic DPO6D aggregate beyond structured supervised fine-tuning, with human corroboration on cognitive depth: DPO-Balanced exceeds Full-SIO-SFT on Carkhuff-H ( $\Delta = +0.637$ ,  $d_z = 1.004$ ,  $p_{\text{Holm}} = 0.0032$ ), although not on Cpsy-H. Second, directional weighting induces reliably distinguishable support profiles. Human annotators rate DPO-Safe higher than DPO-Info on PE-H ( $\Delta = +0.487$ ,  $d_z = 0.800$ ,  $p_{\text{Holm}} = 0.0177$ ) and Carkhuff-H ( $\Delta = +0.736$ ,  $d_z = 1.527$ ,  $p_{\text{Holm}} = 0.0020$ ), whereas DPO-Info is higher on PA-O-H than both DPO-Safe ( $\Delta = +0.961$ ,  $d_z = 1.062$ ,  $p_{\text{Holm}} = 0.0027$ ) and Full-SIO-SFT ( $\Delta = +0.361$ ,  $d_z = 0.773$ ,  $p_{\text{Holm}} = 0.0228$ ). Thus, multi-objective preference construction reshapes how structured support competence is expressed, rather than guaranteeing uniform improvement on every support dimension.

## 6 Additional Analysis

**Cognitive depth.** The gains of structured supervision extend beyond supportive surface form. As shown in Table 5, Full-SIO-SFT increases Carkhuff

| Model               | Carkhuff     | Feeling      | Meaning      |
|---------------------|--------------|--------------|--------------|
| Rewrite-SFT         | 2.662        | 50.6%        | 28.6%        |
| Base + S/I/O Prompt | 3.069        | 58.4%        | 49.3%        |
| Full-SIO-SFT        | 3.196        | 72.6%        | 56.2%        |
| DPO-Balanced        | <b>3.648</b> | <b>91.3%</b> | 78.5%        |
| DPO-Meaning         | 3.626        | 88.1%        | <b>80.8%</b> |
| DPO-Info            | 3.119        | 68.5%        | 54.3%        |

**Table 5:** Cognitive-depth diagnostics derived from the Carkhuff-based evaluation protocol.

| Parameter-space Diagnostic          | Value               |
|-------------------------------------|---------------------|
| $\cos(\Delta_S, \Delta_I)$          | $0.3535 \pm 0.0731$ |
| $\cos(\Delta_S, \Delta_O)$          | $0.3817 \pm 0.0779$ |
| $\cos(\Delta_I, \Delta_O)$          | $0.3631 \pm 0.0708$ |
| Simplex reconstruction of Full-SIO  | 1.8%                |
| Unconstrained linear reconstruction | 14.0%               |

**Table 6:** Parameter-space summary for the S/I/O LoRA updates.

depth over Rewrite-SFT and nearly doubles Meaning recognition (56.2% vs. 28.6%); its dialogue-level Carkhuff gain is significant ( $\Delta = +0.491$ ,  $p_{\text{Holm}} = 0.0012$ ). Prompting recovers part of this benefit, but remains below structured training on both Feeling and Meaning recognition. Preference optimization further differentiates objectives: DPO-Balanced achieves the highest overall depth and Feeling recognition, whereas DPO-Meaning achieves the highest Meaning recognition. DPO-Info, despite greater Objective adherence than the other DPO variants, falls below Full-SIO-SFT on cognitive depth.

**Parameter-space composition.** The behavioral distinction among S/I/O channels is also reflected, although only partially, in adapter space. Table 6 summarizes the central geometric evidence.

The moderate cosine similarities indicate that S/I/O share a common emotional-support foundation without collapsing into a single update direction. At the same time, the low Full-SIO reconstruction ratios suggest that joint structured training is not well approximated by adding independently trained view-specific adapters. Behaviorally, DeltaSVD compositions preserve the expected directional contrast: DeltaSVD-SI exceeds DeltaSVD-IO on PA-S ( $\Delta = +0.196$ ,  $d_z = 0.564$ ,  $p_{\text{Holm}} = 0.0210$ ), whereas DeltaSVD-IO exceeds DeltaSVD-SI on PA-O ( $\Delta = -0.191$  for SI minus IO,  $d_z = -0.647$ ,  $p_{\text{Holm}} = 0.0210$ ). However, the absolute effects remain much weaker than joint training: DeltaSVD-SI improves PA-S by only +0.51 over Base compared with +1.15 for Full-SIO-SFT, and DeltaSVD-IO improves PA-O by only +0.21 compared with +0.82 for Full-SIO-SFT. The resulting claim is therefore deliberately qualified: perspective directionality is partially

| Model        | PE-H        | Carkhuff-H  | Cpsy-H      | PA-O-H      |
|--------------|-------------|-------------|-------------|-------------|
| Rewrite-SFT  | 3.10        | 2.39        | 3.17        | 2.07        |
| Full-SIO-SFT | 3.56        | 2.87        | 3.63        | 2.37        |
| DPO-Balanced | 3.78        | 3.41        | <b>3.72</b> | 2.03        |
| DPO-Safe     | <b>3.91</b> | <b>3.53</b> | 3.64        | 1.84        |
| DPO-Info     | 3.43        | 2.81        | 3.63        | <b>2.66</b> |

**Table 7:** Blinded comparative human ratings from two psychology-background annotators over 50 stratified evaluation instances. Cpsy-H is a simplified four-item human rubric; PA-O-H is a diagnostic of objective-support realization.

compositional in parameter space, but joint Full-SIO training learns stronger cross-perspective behavior than post-hoc adapter merging can recover.

**Blinded human corroboration and evaluation triangulation.** Table 7 reports blinded comparative ratings from two annotators with psychology backgrounds. The human results independently reproduce the central structured-supervision finding: Full-SIO-SFT receives higher descriptive ratings than Rewrite-SFT on every displayed human outcome and significantly exceeds it on PE-H, Carkhuff-H, and Cpsy-H after correction for ten targeted human contrasts. They also sharpen the DPO interpretation. DPO-Balanced significantly exceeds Full-SIO-SFT on human-rated cognitive depth but not on Cpsy-H, while DPO-Safe and DPO-Info exhibit a clear relational-objective profile shift: Safe is higher on PE-H and Carkhuff-H, whereas Info is higher on PA-O-H than both Safe and Full-SIO-SFT.

Annotator agreement and human-judge associations are reported in Appendix M. Human-judge rank association is strongest for PE-H ( $\rho = 0.691$ ) and Carkhuff-H ( $\rho = 0.654$ ), while the more moderate PA-O-H association ( $\rho = 0.518$ ) motivates our conservative use of PA as a behavioral diagnostic. Because Cpsy-H and automatic Cpsy use non-equivalent score formats, their correlation is treated only as exploratory rank agreement.

A reduced Gemma experiment shows the same descriptive ordering as the main Qwen setting: Gemma-Full-SIO-SFT exceeds Gemma-Rewrite-SFT on all four reported automatic outcomes, including DPO6D (3.59 vs. 3.38) and Carkhuff depth (3.24 vs. 2.84). Because this reduced experiment does not include the full DPO analysis or separate confirmatory inference, we treat it as supporting evidence rather than a confirmatory cross-architecture generalization result. Finally, response length is only weakly correlated with evaluation score ( $r = 0.20$ ), reducing the likelihood that

improvements are primarily driven by verbosity. Together, the validation and robustness evidence supports the interpretation that structured supervision improves support behavior, while DPO further improves the aggregate objective and reshapes its expressed profile.

## 7 Discussion

**Structured supervision improves beyond matched rewriting.** The comparison among Rewrite-SFT, Base + S/I/O Prompt, and Full-SIO-SFT clarifies the empirical contribution of structured supervision. Rewrite-SFT shows that matched synthetic rewriting already improves emotional support, and prompting shows that explicit structure is useful even without parameter updates. Under the matched-control design, Full-SIO-SFT nevertheless remains stronger across all four primary outcomes and displays its clearest additional gain on Subjective adherence. This pattern supports S/I/O-conditioned supervision as a substantive training signal rather than a by-product of stronger teacher generation.

**Joint training and selective composition serve different roles.** Full-SIO is the strongest general supervised condition, while SI and IO candidates specialize in different dimensions at inference time. These findings are not contradictory. Joint supervision may establish broad competence across multiple support channels, whereas a particular turn can still benefit from emphasizing a narrower relational or informational composition. This distinction motivates future ESC systems that learn broad structured support representations while selecting their expression according to local conversational need.

**Preference optimization improves and reshapes support profiles.** The DPO results separate aggregate improvement from directional profile control. Approximately uniform Balanced DPO yields a significant automatic DPO6D improvement beyond Full-SIO-SFT, and blinded human ratings independently corroborate an increase in cognitive depth, although not in simplified counseling norms. Directional objectives produce an even clearer human-visible distinction: DPO-Safe is rated higher than DPO-Info on perceived empathy and cognitive depth, whereas DPO-Info is rated higher on objective-support realization than both DPO-Safe and Full-SIO-SFT. Because these shifts do not constitute uniform improvement across all

outcomes, we interpret directional DPO as profile redistribution rather than Pareto improvement. Together with the adapter-space analysis, these results suggest that ESC alignment benefits from both joint learning of support structure and explicit control over which support qualities are emphasized.

## 8 Conclusion

We introduced structured empathic supervision for emotional support conversation, separating Feeling and Meaning understanding targets from Subjective, Interpersonal, and Objective response perspectives. Full-SIO-SFT improves beyond decoding- and budget-matched rewriting and prompt-time structure, while perspective analyses reveal systematic relational-informational specialization. DPO further improves the DPO6D aggregate beyond structured SFT and induces interpretable variation in the support profiles realized by different preference objectives. Together, these results suggest that effective ESC alignment requires both structured empathic supervision and explicit control over competing support objectives.

## 9 Limitations

Our main evaluation is designed as a coverage-oriented diagnostic suite rather than a frequency-weighted estimate of the natural ESConv test distribution. The 20 selected dialogues are chosen using dataset-intrinsic case attributes only, without consulting model generations or evaluation scores, and cover all 11 problem types represented in ESConv. The resulting comparisons therefore characterize performance across this deliberately broad-coverage suite rather than natural-prevalence-weighted test performance.

Target construction and the primary automatic evaluation use the same Claude Sonnet model family. However, the central Full-SIO-SFT versus Rewrite-SFT comparison is reproduced by two non-Claude judges and by blinded human ratings. This reduces, but does not eliminate, concerns about evaluator dependence: automatic scores remain sensitive to rubric design and model-specific judgment. The response-quality study is also limited to two annotators and 18 independent dialogues for inference, and its human perspective-adherence evaluation covers only the Objective channel.

Feeling, Meaning, and S/I/O annotations are operational training targets rather than verified descriptions of users' internal states. Human and

independent-model audits found high overall acceptability, but the human audit covers only 60 turns and mild over-inference remains the main observed error.

Finally, our experiments are limited to English ESConv and 5–25-word supportive responses. The findings may not transfer to longer interactions, other languages or cultures, or clinical and crisis-support settings. We also study only four manually specified preference objectives and LoRA-based parameter composition.

## 10 Ethical Considerations

Emotional-support systems may interact with users experiencing distress or interpersonal difficulty. The models studied here are evaluated as research systems for supportive response generation; they are not designed for diagnosis, crisis intervention, or autonomous mental-health support. Structured supervision may improve response quality but cannot guarantee safety: generated responses may still misunderstand users, overstate empathy, or provide unsuitable guidance.

Feeling and Meaning annotations are computational training signals rather than verified descriptions of users' internal states. Any real-world application should avoid presenting such inferences as facts, communicate the limits of automated support, and provide escalation pathways for high-risk situations. Our claims are restricted to behavior under the stated evaluation framework and do not establish therapeutic benefit or deployment safety.

## References

- Nitay Calderon, Roi Reichart, and Rotem Dror. 2025. [The alternative annotator test for LLM-as-a-judge: How to statistically justify replacing human annotators with LLMs](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16051–16081, Vienna, Austria. Association for Computational Linguistics.
- Yirong Chen, Xiaofen Xing, Jingkai Lin, Huimin Zheng, Zhenyu Wang, Qi Liu, and Xiangmin Xu. 2023. [SoulChat: Improving LLMs' empathy, listening, and comfort abilities through fine-tuning with multi-turn empathy conversations](#). In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 1170–1183, Singapore. Association for Computational Linguistics.
- Arthur J. Clark. 2007. *Empathy in Counseling and Psychotherapy: Perspectives and Practices*. Routledge, New York.

- Arthur J. Clark. 2010. [Empathy: An integral model in the counseling process](#). *Journal of Counseling & Development*, 88(3):348–356.
- Shauna Concannon and Marcus Tomalin. 2024. [Measuring perceived empathy in dialogue systems](#). *AI & Society*, 39(5):2233–2247.
- R. R. Davidson. 1970. [On extending the bradley-terry model to accommodate ties in paired comparison experiments](#). *Journal of the American Statistical Association*, 65(329):317–328.
- Gemma Team. 2025. [Gemma 3 technical report](#). *arXiv preprint arXiv:2503.19786*.
- Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. [LoRA: Low-rank adaptation of large language models](#). In *International Conference on Learning Representations*.
- Dongjin Kang, Sunghwan Kim, Taeyoon Kwon, Seungjun Moon, Hyunsouk Cho, Youngjae Yu, Dongha Lee, and Jinyoung Yeo. 2024. [Can large language models be good emotional supporter? mitigating preference bias on emotional support conversation](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15232–15261, Bangkok, Thailand. Association for Computational Linguistics.
- Aakriti Kumar, Nalin Pongpeth, Diyi Yang, Erina Farrell, Bruce L. Lambert, and Matthew Groh. 2026. [When large language models are reliable for judging empathic communication](#). *Nature Machine Intelligence*, 8:173–185.
- Allison Lahnlala, Charles Welch, David Jurgens, and Lucie Flek. 2025. [The practical impacts of theoretical constructs on empathy modeling](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 15559–15586, Suzhou, China. Association for Computational Linguistics.
- Dawei Li, Bohan Jiang, Liangjie Huang, Alimohammad Beigi, Chengshuai Zhao, Zhen Tan, Amrita Bhat-tacharjee, Yuxuan Jiang, Canyu Chen, Tianhao Wu, Kai Shu, Lu Cheng, and Huan Liu. 2025. [From generation to judgment: Opportunities and challenges of LLM-as-a-judge](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 2757–2791, Suzhou, China. Association for Computational Linguistics.
- Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021. [Towards emotional support dialog systems](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3469–3483, Online. Association for Computational Linguistics.
- Yushan Qian, Weinan Zhang, and Ting Liu. 2023. [Harnessing the power of large language models for empathetic response generation: Empirical investigations and improvements](#). In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 6516–6528, Singapore. Association for Computational Linguistics.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. 2023. [Direct preference optimization: Your language model is secretly a reward model](#). In *Advances in Neural Information Processing Systems*, volume 36.
- Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. [Towards empathetic open-domain conversation models: A new benchmark and dataset](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381, Florence, Italy. Association for Computational Linguistics.
- Jiashuo Wang, Chunpu Xu, Chak Tou Leong, Wenjie Li, and Jing Li. 2024. [Muffin: Mitigating unhelpfulness in emotional support conversations with multifaceted AI feedback](#). In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 567–585, Bangkok, Thailand. Association for Computational Linguistics.
- 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, and 41 others. 2025. [Qwen3 technical report](#). *arXiv preprint arXiv:2505.09388*.
- Yidan Yin, Nan Jia, and Cheryl J. Waksalak. 2024. [AI can help people feel heard, but an AI label diminishes this impact](#). *Proceedings of the National Academy of Sciences*, 121(14):e2319112121.
- Jiahao Yuan, Zhiqing Cui, Hanqing Wang, Yuansheng Gao, Yucheng Zhou, and Usman Naseem. 2025a. [Kardia-R1: Unleashing LLMs to reason toward understanding and empathy for emotional support via rubric-as-judge reinforcement learning](#). *arXiv preprint arXiv:2512.01282*.
- Jiahao Yuan, Zixiang Di, Zhiqing Cui, Guisong Yang, and Usman Naseem. 2025b. [ReflectDiffu: Reflect between emotion-intent contagion and mimicry for empathetic response generation via a RL-diffusion framework](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 25435–25449, Vienna, Austria. Association for Computational Linguistics.
- Chao Zhang, Xin Shi, Xueqiao Zhang, Yifan Zhu, Yi Yang, and Yawei Luo. 2025a. [DecoupledESC: Enhancing emotional support generation via strategy-response decoupled preference optimization](#). In *Findings of the Association for Computational Linguistics*:

EMNLP 2025, pages 22189–22215, Suzhou, China. Association for Computational Linguistics.

Chenhao Zhang, Renhao Li, Minghuan Tan, Min Yang, Jingwei Zhu, Di Yang, Jiahao Zhao, Guancheng Ye, Chengming Li, and Xiping Hu. 2024a. [CPsyCoun: A report-based multi-turn dialogue reconstruction and evaluation framework for Chinese psychological counseling](#). In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 13947–13966, Bangkok, Thailand. Association for Computational Linguistics.

Tenggan Zhang, Xinjie Zhang, Jinming Zhao, Li Zhou, and Qin Jin. 2024b. [ESCoT: Towards interpretable emotional support dialogue systems](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13395–13412, Bangkok, Thailand. Association for Computational Linguistics.

Xinjie Zhang, Wenxuan Wang, and Qin Jin. 2025b. [IntentionESC: An intention-centered framework for enhancing emotional support in dialogue systems](#). In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 26494–26516, Vienna, Austria. Association for Computational Linguistics.

Haiquan Zhao, Lingyu Li, Shisong Chen, Shuqi Kong, Jiaan Wang, Kexin Huang, Tianle Gu, Yixu Wang, Jian Wang, Liang Dandan, Zhixu Li, Yan Teng, Yanghua Xiao, and Yingchun Wang. 2024. [ESC-eval: Evaluating emotion support conversations in large language models](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 15785–15810, Miami, Florida, USA. Association for Computational Linguistics.

## A Structured Annotation and Target Construction Prompts

This section provides the structured annotation and response-target templates used in our pipeline. Full-SIO and Rewrite targets were generated from the same source instances using the same teacher model, explicit decoding configuration, response-length constraint, target format, and supervised-training budget. They differ in whether S/I/O perspective plans and S/I/O-specific integration guidance are included in the target-construction prompt. Verified generation configurations are reported in Appendix B.

### A.1 Feeling and Meaning Annotation

```
You are an expert psychological annotator. Given a dialogue context and the current user utterance, annotate:

1. Feeling (F): The user's immediate emotional state. Use specific emotion labels (e.g., "anxiety about job security", "grief from loss", "frustration with academic pressure"). Avoid vague labels like "sad" or "bad".

2. Meaning (M): The deeper personal significance behind the feeling. Why does this situation matter to this specific user? What value, identity, or relationship is at stake?

Output JSON:
{"feeling": "...", "meaning": "..."}

```

### A.2 S/I/O Perspective-plan Annotation

```
You are an expert in Arthur J. Clark's Integrative Empathy framework (Clark, 2010). Given the dialogue, the user's Feeling and Meaning, generate three perspective plans:

1. SE_plan (Subjective Empathy): How should the supporter momentarily identify with the user's internal frame of reference? What affective states and bodily sensations should the supporter acknowledge?

2. IE_plan (Interpersonal Empathy): How should the supporter validate the user's experience on an immediate here-and-now basis? What relational dynamic or interactional validation is needed?

3. OE_plan (Objective Empathy): What external frame of reference, situational grounding, or practical orientation is appropriate without becoming overly directive?

Output JSON:
{"SE_plan": "...", "IE_plan": "...", "OE_plan": "..."}

```

### A.3 Full-SIO Target Construction

```
System:
You are an emotional support conversationalist trained in Arthur J. Clark's Integrative Empathy framework. Your task is to generate a supportive response that integrates Subjective (S), Interpersonal (I), and Objective (O) empathy perspectives.

User:
<Context>{dialogue_context}</Context>

<CurrentUserText>{user_text}</CurrentUserText>

<Feeling>{feeling}</Feeling>

```

```
<Meaning>{meaning}</Meaning>

<Strategy>{strategy}</Strategy>

<PerspectivePlans>
[SE_plan]: {se_plan}
[IE_plan]: {ie_plan}
[OE_plan]: {oe_plan}
</PerspectivePlans>

Generate a supporter response (5-25 words) that integrates all three perspectives.
First think about how to address each plan, then write the response.

```

### A.4 Rewrite-SFT Target Construction

```
System:
You are an emotional support conversationalist. Generate a supportive response.

User:
<Context>{dialogue_context}</Context>

<CurrentUserText>{user_text}</CurrentUserText>

<Feeling>{feeling}</Feeling>
<Meaning>{meaning}</Meaning>

<Strategy>{strategy}</Strategy>

Generate a supportive response (5-25 words).
First think about the user's emotional state and what support they need, then write the response.

```

Rewrite-SFT uses the same source instances, teacher model, explicit decoding configuration, strategy information, target format, response-length constraint, and supervised-training budget as Full-SIO-SFT, but removes S/I/O-specific planning and integration guidance. It therefore provides the principal matched control for estimating the contribution of S/I/O-conditioned structured supervision.

### A.5 Single-view Target Construction

The single-view prompts follow the Full-SIO template but expose only one perspective plan. The S-only, I-only, and O-only variants differ only in the exposed plan and in the instruction specifying the target perspective.

### A.6 Prompt-only Structured Control

```
System:
You are an emotional support conversationalist trained in Arthur J. Clark's Integrative Empathy framework. Your task is to generate a supportive response that integrates Subjective (S), Interpersonal (I), and Objective (O) empathy perspectives.

User:
<Context>{dialogue_context}</Context>

<CurrentUserText>{user_text}</CurrentUserText>

<Feeling>{feeling}</Feeling>
<Meaning>{meaning}</Meaning>

<Strategy>{strategy}</Strategy>

<PerspectivePlans>
[SE_plan]: {se_plan}
[IE_plan]: {ie_plan}
[OE_plan]: {oe_plan}

```

</PerspectivePlans>

Generate a supportive response (5-25 words).

Unlike Full-SIO-SFT, the prompt-only control supplies the structured representation to the unfine-tuned base model only at inference time.

## B Training and Generation Configuration

### B.1 Annotation and Response-generation Configuration

Table 8 summarizes the verified models used for annotation, target construction, preference construction, and final evaluation.

| Component                    | Model                     |
|------------------------------|---------------------------|
| Feeling / Meaning annotation | GPT-5.2                   |
| S/I/O plan annotation        | GPT-5.2                   |
| Full-SIO targets             | Claude 4.5 Sonnet         |
| Rewrite targets              | Claude 4.5 Sonnet         |
| Prompt-only control          | Qwen3-4B base             |
| Controlled candidates        | Full-SIO-SFT (Qwen3-4B)   |
| Preference judge             | Gemini-3-Flash            |
| Primary evaluator            | Claude 4.5 Sonnet         |
| Robustness evaluators        | GPT-5.2; Gemini-3.5-Flash |

**Table 8:** Verified annotation, generation, preference-judging, and final-evaluation model configurations.

Full-SIO and Rewrite targets use the same source instances, teacher model, explicit decoding configuration, response-length constraint, retry logic, target format, and supervised-training budget; they differ only in whether S/I/O perspective-plan conditioning is supplied. Controlled SI/SO/IO/Full-SIO candidates are generated by the resulting Full-SIO-SFT checkpoint rather than by the teacher model. Preference construction uses Gemini-3-Flash for pairwise judging. Claude 4.5 Sonnet is the primary evaluator, while GPT-5.2 and Gemini-3.5-Flash are used for evaluator-robustness checks. For both target-generation conditions, only the stored target fields used for supervised training are retained; no hidden intermediate reasoning text is used as a training target.

## C Preference Construction Details

### C.1 Controlled Candidate Space

For each of 3,440 training instances, Full-SIO-SFT generates four perspective-conditioned candidate responses:

$$\mathcal{Y}(x) = \{y^{SI}, y^{SO}, y^{IO}, y^{\text{Full-SIO}}\}.$$

All candidates share the same dialogue context, Feeling and Meaning annotations, support strategy,

generator checkpoint, and decoding setting; they differ only in the supplied perspective plans.

### C.2 Pairwise Judging

For each candidate set, we evaluate all six unordered candidate pairs independently on six support dimensions: empathy, identification, effectiveness, sensitivity, non-directiveness, and informational support. Each pair is evaluated twice with swapped response order to reduce position bias. The preference judge is google/gemini-3-flash, accessed through OpenRouter, and returns one of three outcomes: candidate A preferred, candidate B preferred, or tie.

This process yields 247,680 dimension-level pairwise judgments ( $3,440 \text{ instances} \times 6 \text{ pairs} \times 6 \text{ dimensions} \times 2 \text{ response orders}$ ).

### C.3 Pooled Tie Estimation and Within-instance Davidson Ranking

Candidate worths are defined locally within each training instance: the four perspective-conditioned responses for one dialogue context are compared against one another, but their latent values are not assumed to share an absolute scale with candidates from another context. At the same time, estimating a separate tie parameter from the six unordered comparisons available within a single instance would be unstable. We therefore use a mixed estimation strategy.

For each support dimension  $d$ , we first estimate a single global tie parameter  $\hat{\nu}_d$  from tie outcomes pooled across all training instances in that dimension. Holding  $\hat{\nu}_d$  fixed, we then fit a regularized Bradley–Terry–Davidson model independently for each instance  $x$  over its four candidates. Let  $\pi_{x,d,i}$  and  $\pi_{x,d,j}$  denote the latent worths of candidates  $i$  and  $j$  for instance  $x$  under dimension  $d$ . The model defines:

$$P(i \succ j \mid x, d) = \frac{\pi_{x,d,i}}{\pi_{x,d,i} + \pi_{x,d,j} + 2\hat{\nu}_d\sqrt{\pi_{x,d,i}\pi_{x,d,j}}},$$

and

$$P(i = j \mid x, d) = \frac{2\hat{\nu}_d\sqrt{\pi_{x,d,i}\pi_{x,d,j}}}{\pi_{x,d,i} + \pi_{x,d,j} + 2\hat{\nu}_d\sqrt{\pi_{x,d,i}\pi_{x,d,j}}}.$$

Within each instance–dimension fit, one log-worth parameter is fixed to zero for identifiability. We use L2 regularization  $\lambda = 0.1$  and a pseudo-count of 0.05, following the implemented ranking

procedure. The resulting candidate worths are normalized only within the same instance and support dimension:

$$\tilde{s}_d(y | x) = \frac{\pi_{x,d}(y)}{\sum_{y' \in \mathcal{Y}(x)} \pi_{x,d}(y')}.$$

This construction shares only the dimension-specific propensity for ties across instances; candidate worths and downstream rankings remain instance-specific.

#### C.4 Objective-specific Ranking and Pair Construction

Using the within-instance normalized worths, for each preference objective  $w$  we compute:

$$R_w(y | x) = \sum_{d \in \mathcal{D}} w_d \tilde{s}_d(y | x),$$

using the weights reported in Table 1. Candidates are ranked within each instance according to  $R_w$ .

We construct two possible preference pairs:

- **Pair A:** Top1 candidate versus Bottom candidate, included for every retained instance.
- **Pair B:** Top1 candidate versus Full-SIO, included only when Top1 is not Full-SIO.

Pairs are removed when the chosen-to-rejected response-length ratio exceeds 1.3. We additionally validate the required candidate-format prefix and deduplicate by split identifier and pair type, prioritizing Pair A over Pair B. For each of the four reported objectives, matched-size sampling produces 4,094 preference pairs.

### D Evaluation Rubrics and Judge Configuration

#### D.1 Automatic Judge Configuration

All automatic response-quality evaluations use the same final-evaluation judge configuration shown in Table 9. The model used for final automatic evaluation is distinct from the gemini-3-flash model used for DPO preference-pair construction.

#### D.2 DPO6D Evaluation Rubric

DPO6D evaluates six support-quality dimensions on a 1–5 scale:

1. **Empathy:** Does the supporter demonstrate understanding of the user’s emotional state?

| Parameter                            | Value             |
|--------------------------------------|-------------------|
| Judge model                          | Claude 4.5 Sonnet |
| API provider                         | OpenRouter        |
| Temperature / top- $p$               | 0 / 1             |
| Max output tokens: DPO6D / PE / Cpsy | 700               |
| Max output tokens: Carkhuff / PA     | 1500              |

**Table 9:** Automatic final-evaluation judge configuration.

2. **Identification:** Does the supporter identify specific aspects of the user’s situation?
3. **Effectiveness:** Does the response help the user feel better or make progress?
4. **Sensitivity:** Does the supporter avoid being dismissive, judgmental, or presumptuous?
5. **Non-Directiveness:** Does the supporter avoid imposing solutions or advice?
6. **Informational Support:** Does the response provide useful factual information or practical orientation?

The expected output schema is:

```
{
  "scores": {
    "empathy": 1-5,
    "identification": 1-5,
    "effectiveness": 1-5,
    "sensitivity": 1-5,
    "non_directiveness": 1-5,
    "informational": 1-5
  }
}
```

We report DPO6D Total as the unweighted mean across the six dimensions.

#### D.3 Perceived Empathy Evaluation

Perceived Empathy (PE) evaluates whether a response makes the user feel understood and supported. We score nine dimensions on a 1–7 scale: understood, validated, affirmed, accepted, cared for, seen, emotional support, practical advice, and motivation. We report PE as the mean score across these dimensions.

#### D.4 CpsyCounEvalMatrix Evaluation

Cpsy evaluates counseling-oriented response quality across four criterion families: comprehensiveness, professionalism, authenticity, and safety. The automatic Cpsy evaluation uses a 14-item continuous composite score on a 0–10 scale, with a score granularity of 0.1. The human-validation protocol instead uses a simplified four-criterion binary format aggregated on a 0–4 scale. Because the machine and human formats are non-equivalent, we

report human inter-annotator agreement for Cpsy but do not interpret the direct human–judge score correlation as a comparable-scale validity estimate.

### D.5 Carkhuff Cognitive-depth Evaluation

Carkhuff evaluates whether a response recognizes both the user’s expressed emotion and the personal meaning of the situation. The rubric assigns a 1–5 cognitive-depth score using the reference Feeling and Meaning annotations, where 1 indicates a surface-level paraphrase or generic acknowledgement, 3 indicates accurate Feeling recognition with partial Meaning recognition, and 5 indicates deep and personalized recognition of the user’s underlying Meaning. The judge additionally records evidence for Feeling and Meaning recognition, from which we derive the recognition-rate diagnostics reported in the main paper.

### D.6 Plan-referenced Perspective Adherence

Perspective Adherence (PA) measures whether a generated response realizes the intended S/I/O perspective plans. For each evaluation instance, the judge receives the dialogue context, current user utterance, Feeling, Meaning, and the corresponding SE/IE/OE plans. The judge then independently assigns three 1–5 scores: PA-S for adherence to the Subjective perspective plan, PA-I for adherence to the Interpersonal perspective plan, and PA-O for adherence to the Objective perspective plan. All PA scores reported in this paper use the corrected plan-referenced protocol with shared Base and Full-SIO anchors. Only systems for which corrected PA evaluation was executed are reported with PA values. We use PA as a secondary behavioral diagnostic rather than as an overall response-quality metric.

## E Evaluation Subset Construction

### E.1 Coverage-oriented Selection Procedure

The main evaluation set is designed as a coverage-oriented diagnostic suite for comparing support behaviors across heterogeneous emotional-support scenarios. Its construction uses only attributes of the ESConv test dialogues and does not inspect model outputs or evaluation results.

Selection proceeds in two stages:

1. **Stratified superset construction.** We first construct a 100-dialogue superset from the ESConv test set using problem-type metadata only. All six minority problem types are included, while the remaining slots are allocated

proportionally across the five major types using deterministic case-ID sampling.

2. **Fixed-quota diagnostic selection.** We define fixed problem-type quotas for the final 20-dialogue suite and select dialogues within each quota using dataset-intrinsic case attributes only. No generated responses, automatic scores, or inter-system response differences are used during final selection.

The final suite contains 20 dialogues and 219 turn-level evaluation instances, while covering all 11 problem types represented in ESConv. The suite is intended to support controlled diagnostic comparison across diverse support scenarios rather than to estimate frequency-weighted performance over the natural full-test distribution.

### E.2 Problem-type Coverage

| Problem type          | Count | ESConv % | Suite % |
|-----------------------|-------|----------|---------|
| Ongoing depression    | 4     | 26.2     | 20      |
| Breakup with partner  | 3     | 21.0     | 15      |
| Job crisis            | 3     | 19.5     | 15      |
| Problems with friends | 2     | 14.9     | 10      |
| Academic pressure     | 2     | 11.8     | 10      |
| Sleep problems        | 1     | 2.1      | 5       |
| Procrastination       | 1     | 1.5      | 5       |
| Alcohol abuse         | 1     | 1.0      | 5       |
| Issues with parents   | 1     | 1.0      | 5       |
| Appearance anxiety    | 1     | 0.5      | 5       |
| Conflict with parents | 1     | 0.4      | 5       |

**Table 10:** Problem-type composition of the 20-dialogue diagnostic test suite.

### E.3 Statistical Unit

Although the diagnostic suite contains 219 turn-level responses, the independent sampling unit for inference is the dialogue. All confirmatory tests therefore aggregate paired turn-level differences within each of the 20 dialogues before bootstrap resampling, permutation testing, and effect-size calculation.

## F Complete System Results

Table 11 reports the four primary outcome families for all 25 evaluated systems. All values are reported to three decimal places. Corrected plan-referenced PA results are reported separately in Table 16; Original-SFT and Gemma systems are omitted from PA analyses because corrected PA evaluation was not reported for them.

| Model                              | DPO6D        | PE           | Cpsy         | Carkhuff     |
|------------------------------------|--------------|--------------|--------------|--------------|
| <i>Supervised / controls</i>       |              |              |              |              |
| Base                               | 2.751        | 3.309        | 4.494        | 2.110        |
| Original-SFT                       | 2.413        | 2.559        | 4.347        | 1.406        |
| Rewrite-SFT                        | 3.316        | 3.880        | 4.899        | 2.662        |
| S-only-SFT                         | 3.497        | 4.202        | 5.121        | 2.982        |
| I-only-SFT                         | 3.499        | 4.081        | 5.235        | 3.050        |
| O-only-SFT                         | 3.402        | 4.018        | 5.239        | 2.854        |
| Full-SIO-SFT                       | 3.585        | 4.320        | 5.740        | 3.196        |
| Base + S/I/O Prompt                | 3.241        | 4.256        | 5.176        | 3.069        |
| <i>DPO variants</i>                |              |              |              |              |
| DPO-Balanced                       | 3.829        | 4.681        | <b>6.061</b> | <b>3.648</b> |
| DPO-Meaning                        | 3.736        | 4.586        | 5.915        | 3.626        |
| DPO-Info                           | 3.564        | 4.239        | 5.588        | 3.119        |
| DPO-Safe                           | <b>3.831</b> | <b>4.752</b> | 5.730        | 3.580        |
| <i>Parameter-space composition</i> |              |              |              |              |
| DeltaSVD-SI                        | 3.421        | 3.978        | 4.847        | 3.160        |
| DeltaSVD-SO                        | 3.288        | 3.860        | 4.744        | 3.018        |
| DeltaSVD-IO                        | 3.311        | 3.788        | 4.889        | 3.123        |
| DeltaSVD-SIO-Uniform               | 3.444        | 4.012        | 4.880        | 3.142        |
| DeltaSVD-SIO-Best-Recon            | 3.365        | 3.964        | 4.886        | 3.110        |
| ConflictSVD-SI                     | 3.357        | 3.874        | 4.777        | 3.151        |
| ConflictSVD-SO                     | 3.386        | 3.891        | 4.764        | 3.082        |
| ConflictSVD-IO                     | 3.279        | 3.770        | 4.886        | 3.091        |
| ConflictSVD-SIO                    | 3.357        | 3.956        | 4.855        | 3.132        |
| FactorAvg                          | 3.469        | 4.105        | 5.144        | 2.963        |
| <i>Reduced Gemma replication</i>   |              |              |              |              |
| G3-4B-PT Base                      | 2.406        | 2.441        | 4.397        | 1.393        |
| G3-4B-PT Rewrite                   | 3.381        | 4.170        | 5.226        | 2.840        |
| G3-4B-PT Full-SIO                  | 3.585        | 4.283        | 5.560        | 3.237        |

**Table 11:** Complete primary-outcome results for all 25 evaluated systems. Bold marks the strongest value among the principal Full-SIO/DPO systems.

## G Full Statistical Contrasts

For each contrast, paired turn-level score differences are first aggregated within each of the 20 dialogues. We report the mean paired difference  $\Delta$ , percentile 95% confidence intervals from 10,000 dialogue-cluster bootstrap resamples, paired Cohen’s  $d_z$ , and Holm-corrected sign-flip permutation results across the 14 designated confirmatory contrasts. For the four PA contrasts, exact sign-flip enumeration over all  $2^{20} = 1,048,576$  sign assignments is used.

| Contrast                                            | Metric   | $\Delta$ | 95% CI         | $d_z$  | $p_H$   |
|-----------------------------------------------------|----------|----------|----------------|--------|---------|
| <i>RQ1: Structured supervision</i>                  |          |          |                |        |         |
| SIO-Rewrite                                         | DPO6D    | +268     | [.137, .402]   | .869   | .0036   |
| SIO-Rewrite                                         | PE       | +442     | [.275, .602]   | 1.147  | .0014   |
| SIO-Rewrite                                         | Cpsy     | +851     | [.646, 1.058]  | 1.755  | .0013   |
| SIO-Rewrite                                         | Carkhuff | +491     | [.297, .683]   | 1.099  | .0012   |
| SIO-Orig.                                           | DPO6D    | +1.180   | [1.021, 1.343] | 3.179  | .0011   |
| SIO-Base                                            | DPO6D    | +838     | [.699, .974]   | 2.578  | .0010   |
| <i>RQ3: Preference optimization and calibration</i> |          |          |                |        |         |
| Safe-SIO                                            | DPO6D    | +232     | [.124, .348]   | .885   | .0040   |
| Bal.-SIO                                            | DPO6D    | +244     | [.118, .377]   | .823   | .0040   |
| Safe-Info                                           | DPO6D    | +265     | [.166, .368]   | 1.138  | .0009   |
| Safe-Info                                           | Carkhuff | +468     | [.285, .657]   | 1.068  | .0018   |
| Safe-Info                                           | PA-S     | +755     | [.614, .914]   | 2.167  | < .0001 |
| Safe-Info                                           | PA-O     | -.571    | [-.687, -.441] | -1.983 | < .0001 |
| <i>RQ2: Parameter-space composition</i>             |          |          |                |        |         |
| D-SI-D-IO                                           | PA-S     | +196     | [.052, .344]   | .564   | .0210   |
| D-SI-D-IO                                           | PA-O     | -.191    | [-.312, -.062] | -.647  | .0210   |

**Table 12:** Dialogue-level statistics for the 14 designated confirmatory contrasts.  $p_H$  denotes Holm-corrected  $p$ -values; SIO = Full-SIO-SFT, Bal. = DPO-Balanced, and D-SI/D-IO = DeltaSVD-SI/IO.

## H Non-Claude Judge Robustness

We repeat the central Full-SIO-SFT versus Rewrite-SFT comparison using GPT-5.2 and Gemini-3.5-Flash as alternative judges. Only the judge model is changed; prompts, decoding, parsing, and the dialogue-level testing procedure remain fixed.

| Metric   | Claude | GPT-5.2 | Gemini-3.5 |
|----------|--------|---------|------------|
| DPO6D    | +0.268 | +0.279  | +0.206     |
| PE       | +0.442 | +0.582  | +0.573     |
| Cpsy     | +0.850 | +0.847  | +0.493     |
| Carkhuff | +0.491 | +0.572  | +0.461     |

**Table 13:** Mean Full-SIO-SFT minus Rewrite-SFT differences under three judge families. For each non-Claude judge, all four dialogue-level contrasts remain significant after Holm correction ( $p_{\text{Holm}} \leq 0.003$ ).

The direction of the Full-SIO-SFT advantage is therefore unchanged across all three judge families. This result reduces the likelihood that the central comparison is explained by same-family evaluator preference.

## I External Kardia-R1 Reference

We additionally evaluate Kardia-R1 on the same ESConv diagnostic inputs. Kardia-R1 is a 7B persona-grounded model evaluated zero-shot in our situation-centric setting, so it is not a matched control for Full-SIO-SFT.

| Metric   | GPT-5.2  |        | Gemini-3.5-Flash |        |
|----------|----------|--------|------------------|--------|
|          | Full-SIO | Kardia | Full-SIO         | Kardia |
| DPO6D    | 3.425    | 3.260  | 4.014            | 3.802  |
| PE       | 3.816    | 4.108  | 4.848            | 5.003  |
| Cpsy     | 5.637    | 5.477  | 8.425            | 8.045  |
| Carkhuff | 3.156    | 2.399  | 3.457            | 2.881  |

**Table 14:** Descriptive comparison with Kardia-R1 as an external reference. All cells use 219 turns except GPT-5.2 Carkhuff, which uses 218 because one Kardia item was filtered by the evaluation provider.

Kardia-R1 shows a complementary profile: it receives higher perceived-empathy scores, whereas Full-SIO-SFT receives higher cognitive-depth scores under both non-Claude judges. The comparison is descriptive and should not be interpreted as isolating the effect of structured supervision.

## J Additional Perspective and DPO Results

### J.1 Single-view Perspective Adherence

Each single-view model produces its largest target-aligned gain on its corresponding perspective channel (Table 15).

| Model    | PA-S ( $\Delta$ )   | PA-I ( $\Delta$ )   | PA-O ( $\Delta$ )   |
|----------|---------------------|---------------------|---------------------|
| Base     | 1.67 (–)            | 2.34 (–)            | 1.49 (–)            |
| Rewrite  | 2.27 (+0.60)        | 2.78 (+0.44)        | 2.07 (+0.58)        |
| S-only   | <b>2.67</b> (+1.00) | 2.95 (+0.62)        | 1.99 (+0.50)        |
| I-only   | 2.54 (+0.87)        | <b>3.04</b> (+0.70) | 2.03 (+0.54)        |
| O-only   | 2.54 (+0.87)        | 2.94 (+0.60)        | <b>2.21</b> (+0.73) |
| Full-SIO | 2.82 (+1.15)        | 3.20 (+0.86)        | 2.31 (+0.82)        |

**Table 15:** Corrected plan-referenced PA for the single-view ablations.

Rewrite-SFT improves adherence generally, while S-only, I-only, and O-only induce selective target-aligned realization. Full-SIO-SFT provides the strongest joint coverage across the three perspective channels among supervised systems.

### J.2 Full Corrected Perspective-adherence Results

Table 16 reports all systems evaluated with corrected plan-referenced PA. PA is used as a sec-

ondary diagnostic rather than a primary quality outcome.

Among supervised systems, Full-SIO-SFT provides the strongest joint PA profile across the three perspective channels. DPO-Safe slightly exceeds it on PA-S, but does so while substantially reducing PA-I and PA-O; DPO-Info retains greater PA-O within the DPO family. Parameter-composition systems show weaker absolute perspective realization than jointly trained Full-SIO-SFT, supporting a directional but partial compositionality interpretation.

### J.3 Controlled Candidate-composition Analysis

For each of 3,440 training instances, Full-SIO-SFT generates matched SI, SO, IO, and Full-SIO candidates. Tables 17 and 18 summarize aggregate and stage-specific patterns.

IO wins on informational support in 60.9% of instances, whereas SI leads empathy and identification at 31.6% each, sensitivity at 30.6%, and non-directiveness at 29.9%. Stage-specific patterns are exploratory and are not included in the designated confirmatory contrasts.

### J.4 DPO Objective Profiles

The four DPO variants are initialized from the same Full-SIO-SFT checkpoint and trained on matched-size preference datasets. Tables 19 and 20 separate primary outcomes from perspective diagnostics for single-column readability.

| Model        | PA-S ( $\Delta$ ) | PA-I ( $\Delta$ ) | PA-O ( $\Delta$ ) |
|--------------|-------------------|-------------------|-------------------|
| DPO-Safe     | 2.83 (+1.16)      | 2.83 (+0.49)      | 1.50 (+0.01)      |
| Full-SIO     | 2.82 (+1.15)      | 3.20 (+0.86)      | 2.31 (+0.82)      |
| S-only       | 2.67 (+1.00)      | 2.95 (+0.62)      | 1.99 (+0.50)      |
| DPO-Meaning  | 2.59 (+0.92)      | 2.65 (+0.32)      | 1.64 (+0.16)      |
| Base+Prompt  | 2.55 (+0.88)      | 3.11 (+0.77)      | 2.28 (+0.79)      |
| I-only       | 2.54 (+0.87)      | 3.04 (+0.70)      | 2.03 (+0.54)      |
| O-only       | 2.54 (+0.87)      | 2.94 (+0.60)      | 2.21 (+0.73)      |
| DPO-Balanced | 2.53 (+0.86)      | 2.81 (+0.47)      | 1.68 (+0.19)      |
| Rewrite      | 2.27 (+0.60)      | 2.78 (+0.44)      | 2.07 (+0.58)      |
| C-SVD-SI     | 2.21 (+0.54)      | 2.42 (+0.08)      | 1.52 (+0.03)      |
| D-SVD-SI     | 2.18 (+0.51)      | 2.40 (+0.06)      | 1.53 (+0.04)      |
| D-SVD-UNI    | 2.14 (+0.47)      | 2.45 (+0.11)      | 1.65 (+0.16)      |
| C-SVD-SIO    | 2.11 (+0.44)      | 2.59 (+0.25)      | 1.68 (+0.20)      |
| D-SVD-BEST   | 2.11 (+0.43)      | 2.38 (+0.05)      | 1.64 (+0.16)      |
| C-SVD-SO     | 2.07 (+0.40)      | 2.40 (+0.06)      | 1.67 (+0.18)      |
| FactorAvg    | 2.06 (+0.39)      | 2.66 (+0.32)      | 1.94 (+0.45)      |
| DPO-Info     | 2.05 (+0.38)      | 2.47 (+0.13)      | 2.08 (+0.59)      |
| D-SVD-SO     | 2.03 (+0.36)      | 2.37 (+0.04)      | 1.63 (+0.14)      |
| D-SVD-IO     | 2.02 (+0.35)      | 2.40 (+0.06)      | 1.70 (+0.21)      |
| C-SVD-IO     | 2.01 (+0.34)      | 2.49 (+0.15)      | 1.65 (+0.16)      |
| Base         | 1.67 (−)          | 2.34 (−)          | 1.49 (−)          |

**Table 16:** Corrected plan-referenced PA scores for the 21 evaluated systems. Values in parentheses are gains relative to Base. D-SVD denotes DeltaSVD, C-SVD denotes ConflictSVD, UNI denotes SIO-Uniform, and BEST denotes SIO-Best-Recon.

| View     | Top-1 | Rate  | Tendency      |
|----------|-------|-------|---------------|
| SI       | 1,136 | 33.0% | Relational    |
| Full-SIO | 1,018 | 29.6% | Balanced      |
| SO       | 730   | 21.2% | Mixed         |
| IO       | 556   | 16.2% | Informational |

**Table 17:** Top-1 candidate distribution across 3,440 controlled candidate sets.

| View | Early | Middle | Late  |
|------|-------|--------|-------|
| SI   | 30.3% | 35.2%  | 33.4% |
| IO   | 19.6% | 15.1%  | 14.2% |

**Table 18:** Exploratory Top-1 trends by dialogue stage.

| Model        | DPO6D        | PE           | Cpsy         | Carkhuff     |
|--------------|--------------|--------------|--------------|--------------|
| Full-SIO-SFT | 3.585        | 4.320        | 5.740        | 3.196        |
| DPO-Balanced | 3.829        | 4.681        | <b>6.061</b> | <b>3.648</b> |
| DPO-Meaning  | 3.736        | 4.586        | 5.915        | 3.626        |
| DPO-Info     | 3.564        | 4.239        | 5.588        | 3.119        |
| DPO-Safe     | <b>3.831</b> | <b>4.752</b> | 5.730        | 3.580        |

**Table 19:** Primary outcomes for the four reported DPO objectives and Full-SIO-SFT.

| Model        | Feeling      | Meaning      | PA-S ( $\Delta$ )   | PA-O ( $\Delta$ )   |
|--------------|--------------|--------------|---------------------|---------------------|
| Full-SIO-SFT | 72.6%        | 56.2%        | 2.82 (+1.15)        | 2.31 (+0.82)        |
| DPO-Balanced | <b>91.3%</b> | 78.5%        | 2.53 (+0.86)        | 1.68 (+0.19)        |
| DPO-Meaning  | 88.1%        | <b>80.8%</b> | 2.59 (+0.92)        | 1.64 (+0.16)        |
| DPO-Info     | 68.5%        | 54.3%        | 2.05 (+0.38)        | <b>2.08</b> (+0.59) |
| DPO-Safe     | 88.1%        | 76.3%        | <b>2.83</b> (+1.16) | 1.50 (+0.01)        |

**Table 20:** Carkhuff-derived recognition and perspective diagnostics for DPO objective profiles.

## K Parameter-space Composition Details

### K.1 Analysis Objective

The behavioral analyses show that S/I/O supervision produces distinguishable response tendencies. We further examine whether this structure is visible in LoRA adapter updates learned by the single-view models. All analyzed adapters share Qwen3-4B, LoRA rank  $r = 16$ , scaling coefficient  $\alpha = 32$ , target modules, and supervised training settings. For an adapted linear layer,  $\Delta W = \frac{\alpha}{r} B A$ .

All geometric quantities are computed over the induced functional updates  $\Delta W$ , rather than over non-identifiable LoRA factor vectors. For efficiency, Frobenius inner products are evaluated exactly in factorized form without materializing dense matrices:

$$\langle \Delta W_a, \Delta W_b \rangle_F = \text{tr} \left( A_a^\top B_a^\top B_b A_b \right).$$

Cosine similarities and reconstruction objectives are defined using these functional-update inner products.

### K.2 Update Geometry and Reconstruction

Let  $\Delta_S$ ,  $\Delta_I$ , and  $\Delta_O$  denote the S-only, I-only, and O-only updates. Table ?? reports their cosine similarities and tests reconstruction of the jointly trained Full-SIO update.

The optimal simplex coefficients are  $\lambda_S = 0.376$ ,  $\lambda_I = 0.388$ , and  $\lambda_O = 0.236$ . The low re-

construction ratios suggest that joint Full-SIO training learns cross-perspective interactions rather than simply averaging independent S/I/O directions.

### K.3 Rank-preserving DeltaSVD Composition

We construct executable merged adapters by computing  $\Delta_{\text{merge}} = \lambda_S \Delta_S + \lambda_I \Delta_I + \lambda_O \Delta_O$  and applying rank-16 singular-value truncation,  $\hat{\Delta}_{\text{merge}} = \text{SVD}_{16}(\Delta_{\text{merge}})$ .

| Composition          | $\lambda_S$ | $\lambda_I$ | $\lambda_O$ |
|----------------------|-------------|-------------|-------------|
| DeltaSVD-SI          | 0.600       | 0.400       | 0.000       |
| DeltaSVD-SO          | 0.600       | 0.000       | 0.400       |
| DeltaSVD-IO          | 0.000       | 0.500       | 0.500       |
| DeltaSVD-SIO-Uniform | 0.333       | 0.333       | 0.333       |
| DeltaSVD-SIO-Best    | 0.376       | 0.388       | 0.236       |

**Table 21:** Weights used to construct the DeltaSVD adapter family.

### K.4 Behavioral Evaluation of Composed Adapters

| Model             | PA-S ( $\Delta$ ) | PA-I ( $\Delta$ ) | PA-O ( $\Delta$ ) |
|-------------------|-------------------|-------------------|-------------------|
| Full-SIO-SFT      | 2.82 (+1.15)      | 3.20 (+0.86)      | 2.31 (+0.82)      |
| DeltaSVD-SI       | 2.18 (+0.51)      | 2.40 (+0.06)      | 1.53 (+0.04)      |
| DeltaSVD-SO       | 2.03 (+0.36)      | 2.37 (+0.04)      | 1.63 (+0.14)      |
| DeltaSVD-IO       | 2.02 (+0.35)      | 2.40 (+0.06)      | 1.70 (+0.21)      |
| DeltaSVD-SIO-Uni. | 2.14 (+0.47)      | 2.45 (+0.11)      | 1.65 (+0.16)      |
| DeltaSVD-SIO-Best | 2.11 (+0.43)      | 2.38 (+0.05)      | 1.64 (+0.16)      |
| FactorAvg         | 2.06 (+0.39)      | 2.66 (+0.32)      | 1.94 (+0.45)      |

**Table 22:** Corrected PA for principal parameter-space compositions.

The targeted two-view contrast is reliable: DeltaSVD-SI exceeds DeltaSVD-IO on PA-S ( $\Delta = +0.196$ ,  $d_z = 0.564$ ,  $p_{\text{Holm}} = 0.0210$ ), whereas DeltaSVD-IO exceeds DeltaSVD-SI on PA-O ( $\Delta = -0.191$  for SI minus IO,  $d_z = -0.647$ ,  $p_{\text{Holm}} = 0.0210$ ). However, these effects remain substantially weaker than Full-SIO-SFT.

## L Structured Annotation Audit

Two annotators with psychology backgrounds independently audited the same stratified sample of 60 turns. They rated the Feeling, Meaning, S/I/O plans, strategy, and overall acceptability on a five-point scale. Two models that were not used to generate these annotations independently audited 150 additional items.

Across fields, within-one-point agreement between the two human annotators ranges from 96.7% to 98.3%. Mild over-inference is the most common non-acceptable error tag, especially for Meaning and the Subjective/Objective plans. The audit supports the annotations as useful operational super-

vision signals, but not as verified descriptions of users’ internal states.

| Field                  | Human A | Human B | Claude audit | Gemini audit |
|------------------------|---------|---------|--------------|--------------|
| Feeling accuracy       | 96.7%   | 95.0%   | 95.3%        | 98.7%        |
| Meaning accuracy       | 96.7%   | 95.0%   | 95.3%        | 98.7%        |
| Subjective plan fit    | 100.0%  | 93.3%   | 94.7%        | 99.3%        |
| Interpersonal plan fit | 100.0%  | 98.3%   | 98.7%        | 99.3%        |
| Objective plan fit     | 98.3%   | 91.7%   | 95.3%        | 99.3%        |
| Strategy fit           | 98.3%   | 95.0%   | 98.7%        | 97.3%        |
| Overall acceptability  | 100.0%  | 96.7%   | 96.0%        | 98.7%        |

**Table 23:** Percentage of audit ratings at or above 4 on a five-point scale. Each human annotator rated 60 turns; each model auditor rated 150 items.

## M Blinded Human Comparative Evaluation and Robustness

### M.1 Human-rating Protocol

We conduct a blinded comparative human evaluation on 50 turn-level instances sampled from the 219-turn diagnostic suite using strata defined by dialogue stage and problem type. The sample includes 12 early-stage, 15 middle-stage, and 23 late-stage instances and covers all 11 ESConv problem types. For each instance, two annotators with psychology backgrounds independently rate five anonymized responses generated by Rewrite-SFT, Full-SIO-SFT, DPO-Balanced, DPO-Safe, and DPO-Info. Response order is randomized within each instance, and annotators do not observe system identities. In total, the study contains 2,500 model–rubric ratings.

The human evaluation covers perceived empathy (PE-H), Carkhuff cognitive depth (Carkhuff-H), simplified counseling norms (Cpsy-H), and objective-perspective adherence (PA-O-H). We additionally collect a reduced DPO6D-H diagnostic based only on Empathy and Informational Support; because it does not instantiate the complete six-dimensional automatic rubric, we report it only as an auxiliary human diagnostic. Cpsy-H uses four binary human criteria and is therefore not numerically identical to the automatic 14-item Cpsy composite.

For descriptive summaries, we average ratings across the 50 sampled instances and both annotators. For inferential comparisons, the 50 instances are aggregated within their 18 originating dialogues before paired statistical testing. We report dialogue-level mean differences, 95% dialogue-cluster bootstrap confidence intervals, paired Cohen’s  $d_z$ , and sign-flip permutation results with Holm correction over ten targeted human contrasts. Consequently, the dialogue-weighted means in Table 25 need not exactly equal the instance-level descriptive means

in Table 24.

| Model        | DPO6D-H     | PE-H        | Carkhuff-H  | Cpsy-H      | PA-O-H      |
|--------------|-------------|-------------|-------------|-------------|-------------|
| Rewrite-SFT  | 2.80        | 3.10        | 2.39        | 3.17        | 2.07        |
| Full-SIO-SFT | 3.19        | 3.56        | 2.87        | 3.63        | 2.37        |
| DPO-Balanced | 3.16        | 3.78        | 3.41        | <b>3.72</b> | 2.03        |
| DPO-Safe     | <b>3.22</b> | <b>3.91</b> | <b>3.53</b> | 3.64        | 1.84        |
| DPO-Info     | 3.22        | 3.43        | 2.81        | 3.63        | <b>2.66</b> |

**Table 24:** Mean blinded human ratings across 50 stratified evaluation instances and two annotators. DPO6D-H is a reduced two-dimension human diagnostic; Cpsy-H is a simplified four-item human rubric.

### M.2 Targeted Human Contrasts

| Contrast                              | Metric     | $\Delta$ | 95% CI          | $d_z$ | $p_H$ |
|---------------------------------------|------------|----------|-----------------|-------|-------|
| <i>RQ1: Structured supervision</i>    |            |          |                 |       |       |
| SIO-Rewrite                           | PE-H       | +544     | [+.248, +.842]  | .827  | .0210 |
| SIO-Rewrite                           | Carkhuff-H | +484     | [+.213, +.750]  | .799  | .0212 |
| SIO-Rewrite                           | Cpsy-H     | +588     | [+.262, +.924]  | .796  | .0215 |
| SIO-Rewrite                           | PA-O-H     | +313     | [−.002, +.620]  | .450  | .1574 |
| <i>RQ3: DPO quality corroboration</i> |            |          |                 |       |       |
| Bal.–SIO                              | Carkhuff-H | +637     | [+.347, +.924]  | 1.004 | .0032 |
| Bal.–SIO                              | Cpsy-H     | +049     | [−.190, +.278]  | .094  | .7043 |
| <i>RQ3: Directional profile shift</i> |            |          |                 |       |       |
| Safe–Info                             | PE-H       | +487     | [+.205, +.752]  | .800  | .0177 |
| Safe–Info                             | Carkhuff-H | +736     | [+.516, +.951]  | 1.527 | .0020 |
| Info–Safe                             | PA-O-H     | +961     | [+.574, +1.384] | 1.062 | .0027 |
| Info–SIO                              | PA-O-H     | +361     | [+.153, +.581]  | .773  | .0228 |

**Table 25:** Dialogue-level targeted human contrasts over 18 independent dialogues.  $p_H$  denotes Holm-corrected sign-flip permutation results over the ten reported human contrasts. SIO denotes Full-SIO-SFT and Bal. denotes DPO-Balanced.

Eight of the ten targeted human contrasts remain significant after Holm correction. For RQ1, Full-SIO-SFT receives significantly higher human ratings than Rewrite-SFT on PE-H, Carkhuff-H, and Cpsy-H, while the positive PA-O-H difference is not statistically reliable. For RQ3, DPO-Balanced significantly improves human-rated cognitive depth but not Cpsy-H relative to Full-SIO-SFT. The strongest human evidence concerns directional profile differentiation: DPO-Safe is higher on relationally oriented PE-H and Carkhuff-H, whereas DPO-Info is higher on PA-O-H than both DPO-Safe and Full-SIO-SFT.

### M.3 Human Agreement and Judge Association

| Metric     | $\rho_{HH}$ | ICC  | $\rho_{HJ}$ |
|------------|-------------|------|-------------|
| DPO6D-H    | .553        | .514 | .501        |
| PE-H       | .746        | .674 | .691        |
| Carkhuff-H | .753        | .724 | .654        |
| Cpsy-H     | .753        | .613 | .444*       |
| PA-O-H     | .791        | .758 | .518        |

**Table 26:** Annotator agreement and human–judge rank association. All rows use  $N = 250$ .  $\rho_{HH}$  and  $\rho_{HJ}$  are Spearman correlations; ICC denotes ICC(2,1). The starred Cpsy value compares non-equivalent human and automatic scales and is therefore exploratory.

Inter-annotator agreement is highest for PA-O-H, followed by Carkhuff-H, PE-H, and Cpsy-H; DPO6D-H remains positive but lower. Human–judge association is strongest for PE-H and Carkhuff-H. Together with the targeted contrasts, these results independently corroborate the structured-supervision gain and the relational-objective profile shift induced by directional DPO.

| Model             | DPO6D       | PE          | Cpsy        | Carkhuff    |
|-------------------|-------------|-------------|-------------|-------------|
| G3-4B-PT Base     | 2.41        | 2.44        | 4.40        | 1.39        |
| G3-4B-PT Rewrite  | 3.38        | 4.17        | 5.23        | 2.84        |
| G3-4B-PT Full-SIO | <b>3.59</b> | <b>4.28</b> | <b>5.56</b> | <b>3.24</b> |

**Table 27:** Reduced cross-architecture replication from pre-trained Gemma-3-4B-PT.

All three Gemma variants are fine-tuned from the pretrained `google/gemma-3-4b-pt` checkpoint. Because Gemma DPO variants are not trained, this experiment supports transfer of the structured-supervision result rather than replication of the complete DPO-calibration analysis.