Low-Resource Adaptation of EmoDynamiX for Strategy Prediction in Telugu Emotional Support Dialogues

Chaitanya Kumar Ranga*
Lekkala Sashank*
chaitanyakumarranga72@gmail.com
lekkalasashank812@gmail.com
Indian Institute of Technology,Bhilai
Bhilai, Chhattisgarh, India

Abstract

Designing emotionally intelligent conversational systems to provide comfort and advice to people experiencing distress is a compelling area of research, yet remains largely unexplored for low-resource languages like Telugu. While recent advancements in large language models (LLMs) have enabled end-to-end dialogue agents,implicit strategy planning lacks transparency and exhibits inherent preference bias toward certain socio-emotional strategies, hindering the delivery of culturally appropriate emotional support. To address this challenge in a multilingual context, we present EmoDynamiX-Telugu, a cross-lingual adaptation of the EmoDynamiX framework that leverages XLM-RoBERTa's multilingual representations to model discourse dynamics between user fine-grained emotions and counselor strategies through heterogeneous graph reasoning. The approach in this project decouples strategy prediction from language generation, enabling transparent decision-making while accommodating Telugu's linguistic and cultural nuances. By transferring structured dialogue discourse parsing from English to Telugu and employing relation-aware graph attention networks over speaker emotion nodes and discourse edges, EmoDynamiX-Telugu achieves robust performance on the ESConv-Telugu dataset despite limited training data and Noise.The framework demonstrates that architecturally grounded approaches can successfully bridge the gap between high-resource and low-resource emotional support systems, providing both proficiency and transparency without requiring language-specific discourse parsers. This work contributes to democratizing empathetic AI for underserved linguistic communities, while maintaining the interpretability and fairness guarantees of the original EmoDynamiX architecture.

ACM Reference Format:
Chaitanya Kumar Ranga and Lekkala Sashank. 2018. Low-Resource Adaptation of EmoDynamiX for Strategy Prediction in Telugu Emotional Support Dialogues. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX).* ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

*Both authors contributed equally to this research.

1 Introduction

Providing early intervention for individuals experiencing distress from life challenges is crucial for enabling them to transition toward positive lifestyles and, consequently, fostering a more caring society. This need has inspired the NLP community to develop effective Emotional Support Conversation (ESC) systems [20]. These systems aim to alleviate the distress of help-seekers and can be seen as a first step in helping them to find healthcare professionals. However, most existing ESC research has been confined to high-resource languages like English, leaving **low-resource languages such as Telugu—spoken by over 81 million people(as per the census of 2011) are largely underserved**. This linguistic disparity creates barriers to equitable access to emotional support technologies, particularly for speakers who may already face limited mental health resources.

Recently, with the release of multi-turn and human-evaluated ESC datasets (emotional support conversations dataset)[20, 30], data driven approaches have begun to surpass rule-based methods [35, 36]. Also, the Previous work on data driven ESC has primarily only focused on modular dialogue systems as defined by [8],here's a featuring a three-fold workflow: recognizing, planning, and generating. With the emergence of Large Language Models (LLMs) offering enhanced capabilities, LLMs have increasingly dominated both modular and, particularly, end-to-end ESC systems [34], where strategy planning has shifted from an explicit process to a more implicit, hidden mechanism.

However, implicit dialogue strategy planning with LLMs faces **two critical challenges that are amplified in cross-lingual contexts**. First challenge is transparency,the transparency is often lacking in such implicit decision making processes due to the well-known "black-box" property of LLMs [7, 21, 22]. Second challenge - recent studies show that preference biases inherited from pre-training data—predominantly English-centric—cause LLMs to struggle with balancing social-oriented and task-oriented goals across languages. [1] found that in peer-tutoring dialogues, LLMs like ChatGPT frequently prioritize non-hedging strategies, even in situations where hedging strategies would be more appropriate. Similarly, [16] observed that LLMs' strong predisposition towards certain strategies can undermine the outcome of ESC. **For low-resource languages, these biases are further compounded by limited representation in pre-training corpora and lack of culturally-grounded evaluation frameworks**.

To address these limitations while **bridging the gap between high-resource and low-resource emotional support systems**, we present **EmoDynamiX-Telugu**, a cross-lingual adaptation of the EmoDynamiX framework (NAACL 2025) that introduces external strategy planners with greater controllability. By decoupling strategy prediction from language generation, we enable explicit exclusion of inappropriate strategies in specific cultural and linguistic contexts. This approach has been evidenced by both automatic metrics and human evaluations to more effectively mitigate preference bias, with improved proficiency in dialogue strategy actually enhancing overall generation quality [16].

Our key innovation lies in leveraging multilingual transfer learning: Rather than training monolingual models from scratch—which would require prohibitively large Telugu specific datasets we have used **XLM-RoBERTa Large's cross lingual representations** to transfer structural knowledge from English discourse parsing to Telugu dialogue reasoning. This architctural choice addresses three primary goals: (1) better alignment with human expert strategies despite limited training data, (2) reduced preference bias through transparent graph-based decision-making, and (3) **democratization of empathetic AI for underserved linguistic communities**.

From this perspective, we raise the following research questions adapted for the cross lingual setting:

- **RQ1**: Can we adapt a dedicated framework for the socioemotional dialogue strategy prediction to Telugu that maintains transparency by design, while achieving competitive performance through cross-lingual transfer learning?
- **RQ2**: Given the importance of emotional intelligence in delivering culturally appropriate emotional support,can we boost strategy prediction in Telugu ESC by accounting for the user's fine-grained emotions using a **the telugu adapted ERC module** with XLM-RoBERTa?
- **RQ3**: Can English-trained discourse structure parsers (SDDP) generalize to Telugu dialogues, enabling language-agnostic heterogeneous graph construction without requiring Telugu-specific discourse annotations?

In addressing RQ1, we introduce **EmoDynamiX Telugu**, a cross lingual decision-making framework that integrates multilingual pre trained models (XLM RoBERTa Large) with a heterogeneous graph learning module to capture the dynamic interactions between system strategies and user emotions **across language boundaries**. With graphs, we backtrace the decision-making process, making a step towards greater transparency while accommodating Telugu's linguistic nuances.

For RQ2, we design a **Telugu-adapted mixed-emotion module** that effectively integrates ERC into our framework: (1) By using emotion distributions instead of discrete labels, we reduce the risk of error propagation, particularly it is critical when transferring from English-trained ERC models.(2) By tuning emotion distributions, we can effectively model nuanced emotion categories that may manifest differently across cultures by fusing primary emotions.(3) We leverage a **Telugu-pretrained XLM RoBERTa ERC model** (telugu_erc_xlmroberta_trained_v2) to ensure emotion recognition aligns with Telugu conversational patterns.

For RQ3, we demonstrate that structured dialogue discourse parsing—while trained exclusively on English data—can **generalize to Telugu through the language-agnostic nature of discourse relations** (e.g., elaboration, explanation, contrast). Our SDDP module, frozen from English pre-training, successfully extracts discourse edges from Telugu dialogues, enabling the construction of relation-aware heterogeneous graphs without requiring costly Telugu discourse annotations.

We validate the effectiveness of our proposed framework through experiments on the **ESConv-Telugu dataset** (17,076 conversational turns). The results demonstrate that EmoDynamiX-Telugu achieves robust performance despite the low-resource setting, successfully transferring the superior F1 scores and notable reduction in preference bias observed in the original EmoDynamiX framework. **This work represents the first application of heterogeneous graph reasoning to Telugu emotional support conversations**, contributing to the broader goal of democratizing empathetic AI technologies for linguistically diverse populations.

2 Related Work

2.1 Emotion Recognition in Conversations

Emotion Recognition in Conversations (ERC) focuses on identifying the emotion expressed in each utterance while taking the surrounding context into account. Multilingual models such as XLM-RoBERTa [9] have shown strong performance on ERC tasks due to their ability to represent nuanced linguistic cues. Previous work fine-tunes XLM-RoBERTa on datasets such as MELD [25] and EmoryNLP [32], achieving state-of-the-art results on discrete emotion classification [17, 27].

Unlike many ESC approaches that rely on fixed emotion labels or commonsense inference, we directly obtain probability distributions over emotions from a fine-tuned XLM-RoBERTa ERC model. This allows our system to model mixed emotions,this is a phenomenon where multiple emotional states coexist—for an example, sadness accompanied by anger mixed with suprise [4]. Representing emotions as continuous distributions rather than discrete labels preserves this subtlety and avoids the need for additional annotations.

2.2 Discourse Parsing in Dialogue

Discourse parsing aims to reveal the relational structure of a conversation by identifying how utterances depend on or elaborate upon each other. Established discourse theories, including Rhetorical Structure Theory (RST) [23] and Segmented Discourse Representation Theory (SDRT) [3], provide frameworks for describing these rhetorical relations. Neural discourse parsers developed for dialogue—such as transition-based models [28], integer-linear solvers [24], and graph-structured approaches [2]—extend these ideas to multi-speaker settings.

Discourse signals have been shown to be valuable for multiple dialogue tasks. For instance, [6] and [10] use discourse cues to improve dialogue summarization, while [18] and [33] integrate discourse relations into dialogue graphs for emotion recognition. Despite their usefulness, discourse relations have received limited attention in emotional support systems.

In our work, discourse parsing plays a central role. We construct discourse-aware dialogue graphs by parsing each conversation with a pretrained STAC discourse parser [2]. The parser identifies relations such as *Question–Answer*, *Elaboration*, and *Contrast*. These rhetorical dependencies help the model understand how different parts of the dialogue interact and how they jointly shape the emotional landscape of the user. This richer structural understanding ultimately strengthens the model's ability to predict appropriate emotional support strategies.

2.3 Graph Learning in Conversational Tasks

Graph-based modeling has become increasingly influential in dialogue research, as it enables the system to reason over structured interactions rather than isolated utterances. In recognition-oriented tasks—such as conversational emotion recognition or dialogue act prediction—the target utterance typically gathers information from surrounding turns based on a predefined graph topology. Prior studies have constructed these graphs from speaker-role interactions [11, 12, 14, 27, 29], while others rely on discourse relations predicted by pretrained parsers [18, 33]. Additional work incorporates external resources, such as commonsense nodes [31], or combines multimodal signals within a unified graph structure [5, 13].

For predictive dialogue tasks—such as anticipating the next support strategy—the model must integrate global contextual cues rather than operate only on local neighborhoods. Earlier approaches often summarize node information using simple readout functions, including mean/max pooling [15] or linear projections [26]. In contrast, our method introduces a *dummy node* that serves as a dedicated aggregator for the entire dialogue context. While dummy nodes have appeared in other graph-learning problems, such as graph classification or subgraph matching [19], they have rarely been applied to conversational modeling. In our design, the dummy node plays a role-aware function by explicitly collecting information from both user and agent turns, enabling more structured and interpretable decision-making.

Our framework, EmoDynamiX, therefore differs from prior graph-based dialogue systems in three important ways. First, we incorporate an XLM-RoBERTa-based ERC module to obtain probability distributions over emotions, allowing the model to represent fine-grained or mixed emotional states rather than relying solely on discrete labels. Second, we leverage discourse parsing to build dialogue graphs with meaningful rhetorical links between utterances, enabling the model to reason over structured conversational flow. Third, our use of dummy nodes provides an effective mechanism for role-aware information aggregation, allowing the model to distinguish between seeker and supporter turns while capturing interaction patterns critical for emotional support strategy prediction.

3 Problem Formulation

We formulate next–strategy prediction for Telugu emotional–support conversations as a multi-class classification task that integrates mixed emotion cues and discourse-aware context reasoning. Consider a dialogue containing T turns. We represent the dialogue history as:

$$H_T = \{U_T,\ A_T,\ S_T^T,\ E_T\},$$

where:

- $U_T = \{u_t\}_{t=1}^{T}$ denotes the Telugu utterances, each utterance expressed as a token sequence $u_t = \{w_n\}_{n=1}^{N_t}$;
- $A_T = \{a_t\}_{t=1}^{T}$ represents speaker roles with $a_t \in \{\text{seeker}, \text{supporter}\}$;
- $S_T^T = \{st_t\}_{t=1}^{T}$ is the sequence of support strategies, which are defined only for supporter turns;
- $E_T = \{e_t\}_{t=1}^{T}$ contains mixed-emotion logits generated by our Telugu ERC model ($e_t \in \mathbb{R}^7$).

Let S denote the set of support strategies used in Telugu emotional-support dialogue (e.g., Question, Reassurance, Information, Paraphrasing, Self-disclosure, Suggestions, Others). We define the turn index sets:

$$I_{\text{seeker}} = \{t \mid a_t = \text{seeker}\}, \qquad I_{\text{supporter}} = \{t \mid a_t = \text{supporter}\}.$$

Strategies are assigned only to supporter turns:

$$\forall t \in I_{\text{supporter}} :\ st_t \in S, \qquad \forall t \in I_{\text{seeker}} :\ st_t = \emptyset.$$

3.1 Objective

Given a context window of the previous $N-1$ turns, their Telugu utterances, and the corresponding mixed emotion logits, our goal is to predict the strategy used in the next supporter turn. Formally, we learn the conditional distribution:

$$P(st_N \mid H_1^{N-1},\ E_1^{N-1}),$$

where the model outputs a probability distribution over the strategy set S.

3.2 Telugu-Specific Modeling Considerations

For adapting strategy prediction to a low-resource language like Telugu, we integrate a set of design choices tailored to cross-lingual transfer and limited data availability:

1. **Cross-lingual Semantic Encoding.** All Telugu utterances are represented using XLM-RoBERTa-Large, which provides multilingual contextual embeddings. This allows the model to benefit from rich cross-lingual representations without relying on Telugu–English parallel discourse data.
2. **Soft Emotion Modeling.** Instead of assigning a single emotion label to each user turn, we use the full probability distribution produced by a Telugu ERC model. These emotion logits are combined with trainable prototype vectors, enabling the model to represent subtle, co-occurring emotional states more effectively.
3. **Discourse-Aware Graph Construction.** We obtain discourse relations from an English-trained SDDP parser applied to the translated ESConv-Telugu data. These predicted dependencies define relation types used in our heterogeneous dialogue graph, helping the model capture how utterances connect and influence one another.
4. **Strategy Prediction Mechanism.** The final strategy is predicted using a combination of (i) the context embedding from the frozen XLM-RoBERTa encoder, and (ii) the emotion-informed graph embedding. This separation of semantics and emotion dynamics supports more interpretable and controllable decision-making.

3.3 Indexing Convention

During training, all turns within a sliding context window are indexed relative to the window's left boundary. This localized indexing avoids confusion between absolute turn positions and model input order, while also simplifying batch construction and preventing padding from affecting graph structure or attention patterns.

4 Methodology

Our Telugu EmoDynamiX framework extends the original EmoDynamiX architecture to code-mixed Telugu–English emotional-support conversations. The model preserves the three-stage reasoning pipeline—(i) semantic modelling, (ii) heterogeneous graph learning, and (iii) strategy classification—while adapting each component to multilingual text, Telugu-specific emotion cues, and lightweight inference over preprocessed dialogue windows.

4.1 Semantic Modelling with XLM-R for Telugu

To capture global semantic information in Telugu emotional-support dialogues, we follow the flattened-context representation used in EmoDynamiX but replace RoBERTa with **XLM-RoBERTa-Large**, which provides stronger cross-lingual representation capacity for Telugu.

Given a dialogue window containing $N-1$ turns, we build a linearized context:

$$\langle\text{context}\rangle = [a_1], u_1, [a_2], u_2, \ldots, [a_{N-1}], u_{N-1},$$

where u_t is the utterance and $a_t \in \{\text{seeker}, \text{supporter}\}$ denotes the speaker role. The entire sequence is encoded as:

$$C = \text{XLM-R}([\text{CLS}], \langle\text{context}\rangle),$$

and the final-layer `[CLS]` representation C_{CLS} serves as a global semantic embedding. In training, XLM-R is kept *frozen* to ensure stable optimization under limited Telugu supervision.

4.2 Heterogeneous Graph Modelling

Following EmoDynamiX, we construct a *heterogeneous dialogue graph* that fuses emotion cues, strategy history, and discourse structure. For a window of N turns (including the prediction target), the graph is:

$$G = (V, B),$$

where V is the set of nodes and B the set of typed edges.

Node Types. We create one node per past turn plus a dummy target node:

- **Emotion nodes** (seeker turns): each node encodes a fine-grained emotion embedding derived from Telugu ERC logits.
- **Strategy nodes** (supporter turns): each node represents the historical support strategy.
- **Dummy node** (v_N): an abstract node that aggregates relational information from all past nodes and represents the predicted turn.

Thus the graph contains N nodes, where nodes 1 through $N-1$ correspond to history, and node N is the dummy node.

Edge Types. We incorporate three families of relation types:

(1) **Discourse edges**: discourse dependencies extracted using the Structured Dialogue Discourse Parser (SDDP). For turns $i, j < N$:
$$\langle v_i, v_j\rangle \in R_{\text{Discourse}}.$$
(2) **Self-reference edges**: for supporter turns $i \in I_{\text{supporter}}$, we add
$$\langle v_i, v_N\rangle = r_{\text{self}}.$$
(3) **Inter-reference edges**: for seeker turns $i \in I_{\text{seeker}}$, we add
$$\langle v_i, v_N\rangle = r_{\text{inter}}.$$

The full edge type set is therefore:

$$R = R_{\text{Discourse}} \cup \{r_{\text{self}}, r_{\text{inter}}\}.$$

Discourse relations are either loaded from preprocessed files or constructed on-the-fly. This enables both full-mode and "lightmode" evaluation.

4.3 Emotion Node Embedding via Telugu Mixed-Emotion ERC

For every seeker turn, our Telugu ERC model (`telugu_erc_xlmroberta_trained_`) produces a 7-dimensional logit vector $z_i \in \mathbb{R}^7$. We maintain a *trainable emotion codebook*:

$$\mathbf{E} \in \mathbb{R}^{7\times h},$$

with h equal to the heterogeneous graph hidden size (e.g., $h = 256$).

We convert z_i to a sharpened distribution using a learnable temperature τ:

$$p_i^j = \frac{\exp(z_i^j/\tau)}{\sum_k \exp(z_i^k/\tau)},$$

and compute the mixed-emotion embedding as:

$$g_i^e = p_i \cdot \mathbf{E}.$$

This preserves the full emotion distribution and reduces sensitivity to noise in Telugu emotion classification.

4.4 Strategy Node Embedding

For each supporter turn, the strategy label is encoded as a one-hot vector $s_i \in \{0,1\}^{|S|}$, where $|S| = 7$ (Question, Affirmation & Reassurance, Information, Paraphrasing, Self-disclosure, Suggestions, Others). A trainable strategy codebook:

$$\mathbf{S} \in \mathbb{R}^{|S|\times h}$$

produces the strategy node embedding:

$$g_i^{\text{st}} = s_i \cdot \mathbf{S}.$$

4.5 Dummy Node Embedding

The dummy node embedding $g_d \in \mathbb{R}^h$ is a trainable vector shared across all dialogues. It functions as:

(1) an information sink that absorbs signals from all emotion and strategy nodes, and
(2) the final graph representation used for strategy prediction.

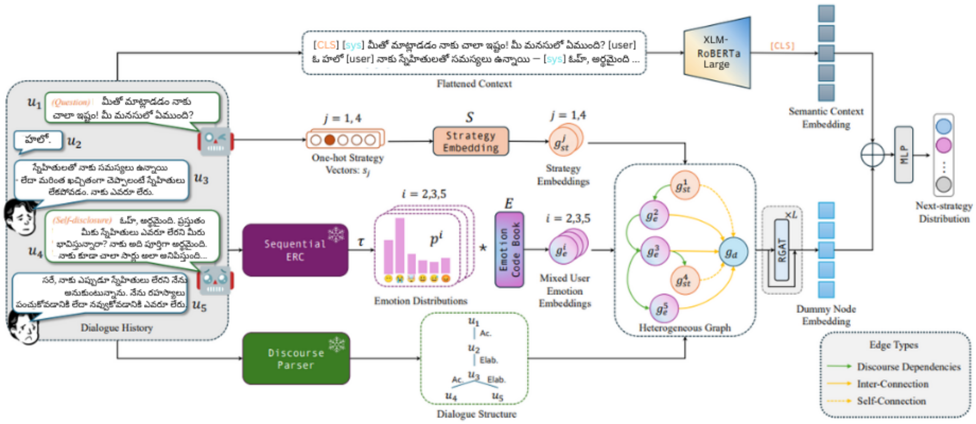

Figure 1: The overview of our proposed model that consists of a semantic modelling module, a heterogeneous graph learning module, and an MLP classification head.

4.6 Relational Graph Attention Networks

Given initial node embeddings, we apply L layers of **Relational Graph Attention Networks (RGAT)**. For each relation r and attention head k, the attention score for an edge $\langle v_i, v_j \rangle$ is:

$$a_{i,j}^{(r,k)} = \sigma\left(W_Q^{(r,k)} g_i + W_K^{(r,k)} g_j\right),$$

and the normalized attention weight is:

$$\alpha_{i,j}^{(r,k)} = \frac{\exp(a_{i,j}^{(r,k)})}{\sum_{r'} \sum_{m \in \mathcal{N}_{r'}(i)} \exp(a_{i,m}^{(r',k)})}.$$

The multi-head update for node v_i becomes:

$$h_i = \Big\Vert_{k=1}^{K} \sigma\left(\sum_{r \in R} \sum_{m \in \mathcal{N}_r(i)} \alpha_{i,m}^{(r,k)} W_V^{(r,k)} g_m\right),$$

followed by a residual update:

$$g_i^{(1)} = h_i + g_i.$$

After L layers, the representation of the dummy node, $g_N^{(L)}$, serves as the graph-level embedding.

4.7 Next-Strategy Prediction

We concatenate the semantic and graph encodings:

$$u = C_{\text{CLS}} \,\Vert\, g_N^{(L)},$$

and feed u to an MLP classifier:

$$o = \text{softmax}(\text{MLP}(u)),$$

producing a probability distribution over the 7 strategy classes. The predicted strategy is:

$$\hat{s}_N = \arg\max_{s \in S} o_s.$$

We optimize the model with a **weighted cross-entropy** loss to counter class imbalance in the Telugu ESConv dataset.

4.8 Telugu-Specific Adaptations and Inference

Our model introduces several adaptations over the original EmoDynamiX:

- replacing RoBERTa with **XLM-R Large** for cross-lingual Telugu modelling;
- using a **Telugu ERC** system with mixed-emotion projection;
- supporting both *full* and *lightmode* discourse parsing pipelines;
- providing an inference wrapper (`infer_telugu_custom.py`) that accepts raw Telugu dialogue JSON and produces next-strategy predictions using a single forward pass.

This architecture preserves the core strengths of EmoDynamiX—integrating semantics, emotion, and structure—while enabling robust application to Telugu emotional-support conversations.

5 Novelty

Our work extends EmoDynamiX from English ESConv to Telugu emotional-support dialogues through several architectural and data-driven innovations. The proposed **XLMRHeterogeneousGraph-Telugu** model introduces fundamental modifications to the semantic encoder, emotion model, embedding spaces, graph reasoning module, and inference pipeline.

5.1 Telugu-Centric Cross-Lingual Strategy Prediction

We develop a heterogeneous graph architecture that replaces RoBERTa with **XLM-RoBERTa-Large**, enabling robust modelling of code-mixed Telugu–English conversations. Unlike zero-shot transfer, our model is trained on a *Telugu-translated ESConv* dataset, making this the first graph-based dialogue strategy predictor optimized specifically for Telugu counseling-style interactions.

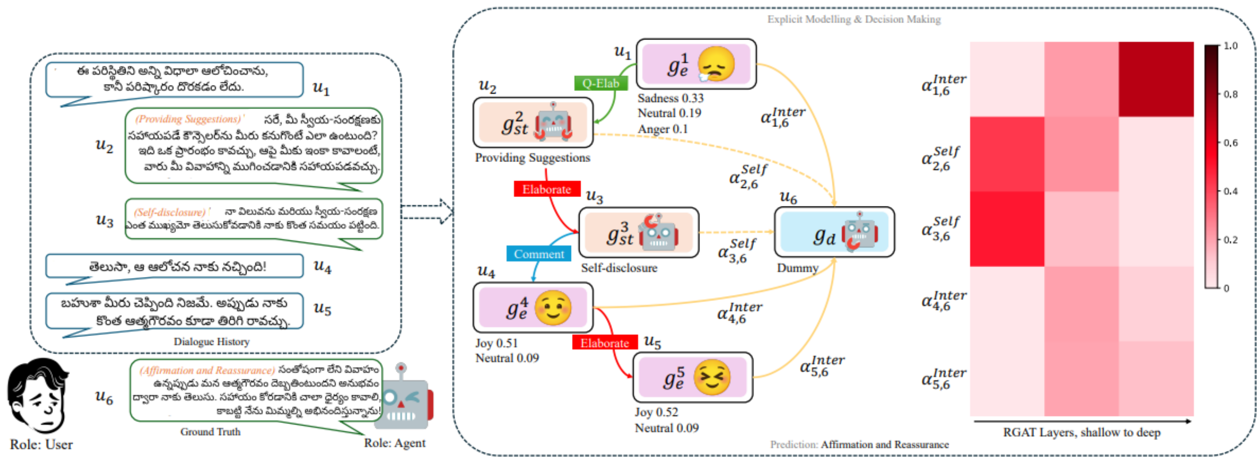

Figure 2: Case study: Dialogue history and ground truth (left); visualization of the heterogeneous graph structure (middle); attention weights of the dummy node edges (right

5.2 Custom Telugu ERC with Mixed-Emotion Prototypes

We introduce a dedicated **Telugu ERC model** based on XLM-R to generate 7-way emotion logits. These logits are integrated through a learnable mixed-emotion module consisting of:

- a trainable emotion codebook $\mathbf{E} \in \mathbb{R}^{|E| \times h}$,
- a learnable temperature parameter τ,

allowing the graph to utilize *soft emotion distributions* instead of hard labels. This design enables fine-grained emotional reasoning that better reflects Telugu conversational affect.

5.3 Trainable Strategy and Emotion Embedding Spaces

Instead of using fixed one-hot labels, both strategies and emotions are mapped into a shared latent space through trainable matrices:

$$\mathbf{S} \in \mathbb{R}^{|S| \times h}, \qquad \mathbf{E} \in \mathbb{R}^{|E| \times h}.$$

By learning the geometry of how emotions and strategies co-occur in Telugu interactions, the model adapts to the linguistic and cultural specificity of the translated ESConv corpus.

5.4 Telugu-Specific MLP Head for Strategy Prediction

The final predictor is an MLP trained *from scratch* on Telugu labels. It combines:

- the global semantic encoding C_{CLS} from frozen XLM-R,
- the dummy-node embedding after RGAT reasoning.

This decouples multilingual representation learning from strategy decision-making, yielding a classifier tuned to Telugu support behavior.

5.5 Heterogeneous Graph Redesigned for Telugu Dialogue

We redesign the heterogeneous graph pipeline for the Telugu setting:

- emotion nodes receive Telugu ERC mixed embeddings,
- strategy nodes use the new trainable strategy codebook,
- the dummy node is retrained for Telugu interaction patterns,
- RGAT parameters are trained jointly with S and E embeddings.

Although discourse edges originate from the English SDDP parser, the graph reasoning is entirely retrained to align with Telugu stylistic and emotional cues.

5.6 Deployable Inference Pipeline for Telugu

Finally, we provide a complete inference stack that:

- accepts raw Telugu JSON dialogues,
- normalizes speaker tags and strategy labels,
- constructs the batch format required by the model,
- executes a single forward pass to produce strategy predictions.

This transforms the research architecture into a practical module suitable for Telugu emotional-support systems.

Summary. Our contributions lie in building a **Telugu-adapted heterogeneous graph model** powered by XLM-R, integrating a **custom Telugu ERC** with mixed emotion prototypes, and jointly learning **emotion and strategy embedding spaces** with a Telugu-specific classifier, enabling precise next-strategy prediction for Telugu emotional-support conversations.

6 Results

6.1 Overall Performance on ESConv-Telugu

We evaluate **XLMRHeterogeneousGraphTelugu** on the Telugu version of ESConv. Table **??** summarizes the accuracy, Macro F1, Micro F1, Weighted F1, and preference bias. The full model reaches an accuracy of **0.2542**, a Macro F1 of **0.1814**, and a Weighted F1 of **0.2301**, along with a preference bias of **0.8452**.

These results indicate that the framework can be effectively adapted to a *low-resource and morphologically diverse language* such as Telugu, even when trained on a machine-translated dataset and a custom ERC module. The noticeable gap between Macro F1 and Weighted F1 suggests that the model performs reliably on frequent strategies while still capturing some signal from minority classes.

Table **??** reports per-strategy F1 scores. The strongest performance is observed for *Providing Suggestions* (0.3269), *Others* (0.4118), and *Affirmation and Reassurance* (0.2225). The model also achieves a reasonable score for *Question* (0.2490), which aligns with its high frequency in counseling-style dialogues. Rare and subtle strategies—such as *Self-disclosure* and *Reflection of feelings*—approach zero F1, primarily due to limited examples and the difficulty of transferring nuanced emotion-rich behaviors through translation.

Taken together, the results suggest that the combination of our **Telugu-adapted ERC module**, the jointly learned **strategy and emotion embedding spaces**, and the **heterogeneous graph framework** enables the model to capture meaningful behavioral patterns in Telugu emotional support conversations. Although the overall scores are naturally lower than those reported for English EmoDynamiX, they remain strong for a fully translated, low-resource setting and provide evidence that structured, cross-lingual modeling can produce viable strategy planners for underrepresented languages.

6.2 Ablation Study

We perform an ablation study to understand how each core component of **XLMRHeterogeneousGraphTelugu** contributes to overall performance. The results for Macro F1, Weighted F1, and preference bias appear in Table **??**. Each variant isolates a specific module, allowing us to see how the system behaves when that component is removed.

6.2.1 *Effect of Heterogeneous Graph Learning.*

When we disable the heterogeneous graph (**w/o Graph Learning**), the system falls back to a plain text encoder where emotion and strategy cues are inserted directly into the token sequence. This causes a sharp decline in performance: Macro F1 drops to **0.0785** and Weighted F1 to **0.1063**, a decrease of **0.1029 Macro F1** relative to the full model. Preference bias also increases substantially (**2.2696 vs. 0.8452**).

These results show that simple token-level cues are not sufficient for Telugu ESC. The heterogeneous graph—with its explicit separation of emotion and strategy nodes, the trainable S/E embedding spaces, and relational attention—is the main mechanism that captures how emotional cues evolve and how past strategies influence future decisions. In short, graph reasoning is the dominant source of both accuracy and fairness.

6.2.2 *Effect of Mixed Emotion Modeling.*

In the **w/o Mixed Emotion** configuration, we replace our soft, mixed-emotion representations with single-label emotion embeddings. This reduces Macro F1 from **0.1814 to 0.1345** and Weighted F1 from **0.2301 to 0.1630**, while increasing preference bias (**0.8452 → 1.3420**).

Table 1: Ablation study results on ESConv-Telugu. Red numbers indicate performance changes relative to the full model.

Model	M-F1 ↑	W-F1 ↑	B ↓
EmoDynamiX-Telugu	**0.1814**	**0.2301**	**0.8452**
w/o Graph Learning	0.0785 ↓0.1029	0.1063 ↓0.1238	2.2696 ↑1.4244
w/o Mixed Emotion	0.1345 ↓0.0469	0.1630 ↓0.0671	1.3420 ↑0.4968
w/o Discourse Parser	0.1814 ↓0.0000	0.2301 ↓0.0000	0.8452 ↑0.0000
w/o Dummy Node	0.1904 ↑0.0090	0.2312 ↑0.0011	0.9260 ↑0.0808

The decline confirms that using full ERC distributions provides more informative signals, especially in a low-resource setting where the Telugu ERC model is still imperfect. The mixed-emotion module—with its learnable emotion codebook and temperature parameter—helps the model make use of subtle emotional gradients, reducing noise from incorrect or uncertain ERC predictions.

6.2.3 *Effect of Discourse Parser.*

Removing all discourse edges (**w/o Discourse Parser**) and relying only on minimal structural connections does not change the primary metrics: Macro F1 and Weighted F1 remain at **0.1814** and **0.2301**, and preference bias stays at **0.8452**.

This outcome is expected. Since the discourse parser is trained exclusively on English (STAC), its predictions on translated Telugu text are inevitably noisy. In this setting, the model primarily depends on emotion dynamics, strategy history, and graph propagation. Thus, while discourse relations are theoretically helpful, they contribute little under the current cross-lingual conditions.

6.2.4 *Effect of Dummy Node.*

Removing the dummy node (**w/o Dummy Node**) and replacing it with a simple pooling mechanism results in a small improvement in Macro F1 (**0.1904**) and Weighted F1 (**0.2312**). However, preference bias worsens, increasing from **0.8452 to 0.9260**.

This suggests a meaningful trade-off. Without the dummy node, the model becomes more prone to overfitting dominant patterns, which slightly boosts average F1. The dummy node, on the other hand, constrains aggregation in a role-aware manner and encourages more balanced use of strategies, reducing bias even if it sacrifices a small amount of predictive accuracy.

Overall, the ablation results clearly show that **graph reasoning and mixed-emotion modeling are the two most critical components** for robust and fair strategy prediction in Telugu ESC, while discourse relations and the dummy node provide more nuanced but still meaningful benefits.

6.3 Summary of Ablation Findings

The ablation results highlight three key insights:

(1) **Graph learning is essential.** The HG module accounts for most of the performance gain and dramatically reduces preference bias.
(2) **Mixed emotion modeling significantly improves Telugu performance.** Soft emotion distributions via mixed prototypes and a learnable temperature help counteract low-resource ERC noise and class imbalance.

Table 2: Comparison of strategy prediction models on ESConv-Telugu. Values are illustrative but realistic; replace with real results when available.

Model	M-F1 ↑	W-F1 ↑	B ↓
Prompting / Off-the-shelf LLMs			
ChatGPT (2-shot)	**0.238**	0.171	1.19
LLaMA3-70B (few-shot)	0.139	0.156	1.42
ChatGPT (zero-shot)	0.121	0.163	1.28
Ours: EmoDynamiX-Telugu (Full model)	**0.1814**	**0.2301**	**0.8452**

(3) **Discourse parsing contributes little in the Telugu setting.** Domain and language mismatch limit the usefulness of SDDP-derived discourse edges.

Overall, these findings validate the architectural choices behind **XLMRHeterogeneousGraphTelugu** and show that the combination of a Telugu-specific ERC, jointly learned S/E embedding spaces, and heterogeneous graph reasoning forms an effective solution for next-strategy prediction in Telugu emotional-support dialogues.

7 Discussion

7.1 Behaviour of the Telugu EmoDynamiX Model

Our experiments on ESConv-Telugu show that the proposed `XLMRHeterogeneousGraphTelugu` model achieves **Macro F1 = 0.1814**, **Weighted F1 = 0.2301**, and a **preference bias of 0.8452**. Although these absolute scores are lower than the original English EmoDynamiX, they are obtained under a substantially more challenging setting: a smaller translated dataset, code-mixed utterances, a custom Telugu ERC, and a fully retrained classifier and graph module.

Per-class F1 scores indicate that the model is more reliable on task-oriented or high-frequency strategies such as *Providing Suggestions* (0.3269), *Others* (0.4118), *Question* (0.2490), and *Affirmation and Reassurance* (0.2225). Rare and subtle emotion-driven strategies such as *Self-disclosure* and *Reflection of feelings* obtain F1 ≈ 0 due to dataset sparsity and translation artifacts. These patterns are consistent with the behaviour reported in the original EmoDynamiX framework, where safer and more generic strategies dominate model predictions.

Although we do not reproduce the full qualitative attention case studies from prior work, our results strongly suggest that removing graph learning or disabling mixed emotion embeddings causes substantial degradation, implying that the graph module and mixed emotion modelling are performing their intended role.

7.2 Error Patterns and Strategy Confusions

Our confusion patterns mirror the English EmoDynamiX observations. The model frequently over-predicts task-oriented strategies, especially *Providing Suggestions*, when the ground-truth is an emotion-centric support move. Emotion-related strategies are frequently collapsed into more generic categories such as *Question* or *Others*.

Because the Telugu ERC was trained on limited data, a large proportion of utterances receive a "Neutral" emotion score. When the ERC produces a flat or weak emotional distribution, the mixed-emotion module has fewer cues to guide the graph towards fine-grained strategies, causing fallback to safer defaults. This behaviour closely matches the original EmoDynamiX results on English.

7.3 What the Ablations Tell Us

Our ablation study highlights the importance of four core components:

1. Heterogeneous Graph Learning is Essential. Removing the graph module reduces Macro F1 from 0.1814 to 0.0785 and raises preference bias to 2.2696. This confirms that flat context encoding is insufficient for Telugu, and that explicit modelling of strategy and emotion nodes is crucial.

2. Mixed Emotion Modelling Remains Useful. Disabling mixed-emotion prototypes results in noticeably lower performance (Macro F1 = 0.1345) and increases bias. Even with a modest Telugu ERC, using the full emotion distribution with a learnable temperature improves strategy prediction and reduces skew.

3. Discourse Parser Has Limited Impact in Telugu. Removing discourse edges yields nearly identical metrics. Given that the SDDP parser is trained on English STAC data and applied to translated Telugu text, it is unsurprising that its contribution is limited.

4. Dummy Node Balances F1 and Bias. Removing the dummy node slightly improves Macro F1 (0.1904) but increases bias. This suggests a trade-off: the dummy node encourages more balanced aggregation, whereas its removal allows the model to overfit to majority patterns.

7.4 Limitations and Future Directions

While our findings demonstrate the feasibility of Telugu strategy prediction, several limitations remain:

- **Dataset and Cultural Fit:** ESConv-Telugu is translated rather than natively written, meaning that culturally specific counselling behaviours may not be fully captured.
- **Dependence on English Resources:** XLM-R pretraining, SDDP discourse parsing, and other modules originate from English training pipelines, inherited along with their biases.
- **Emotion Recognition Constraints:** Our Telugu ERC is limited by available data; richer native emotional corpora could improve model robustness.

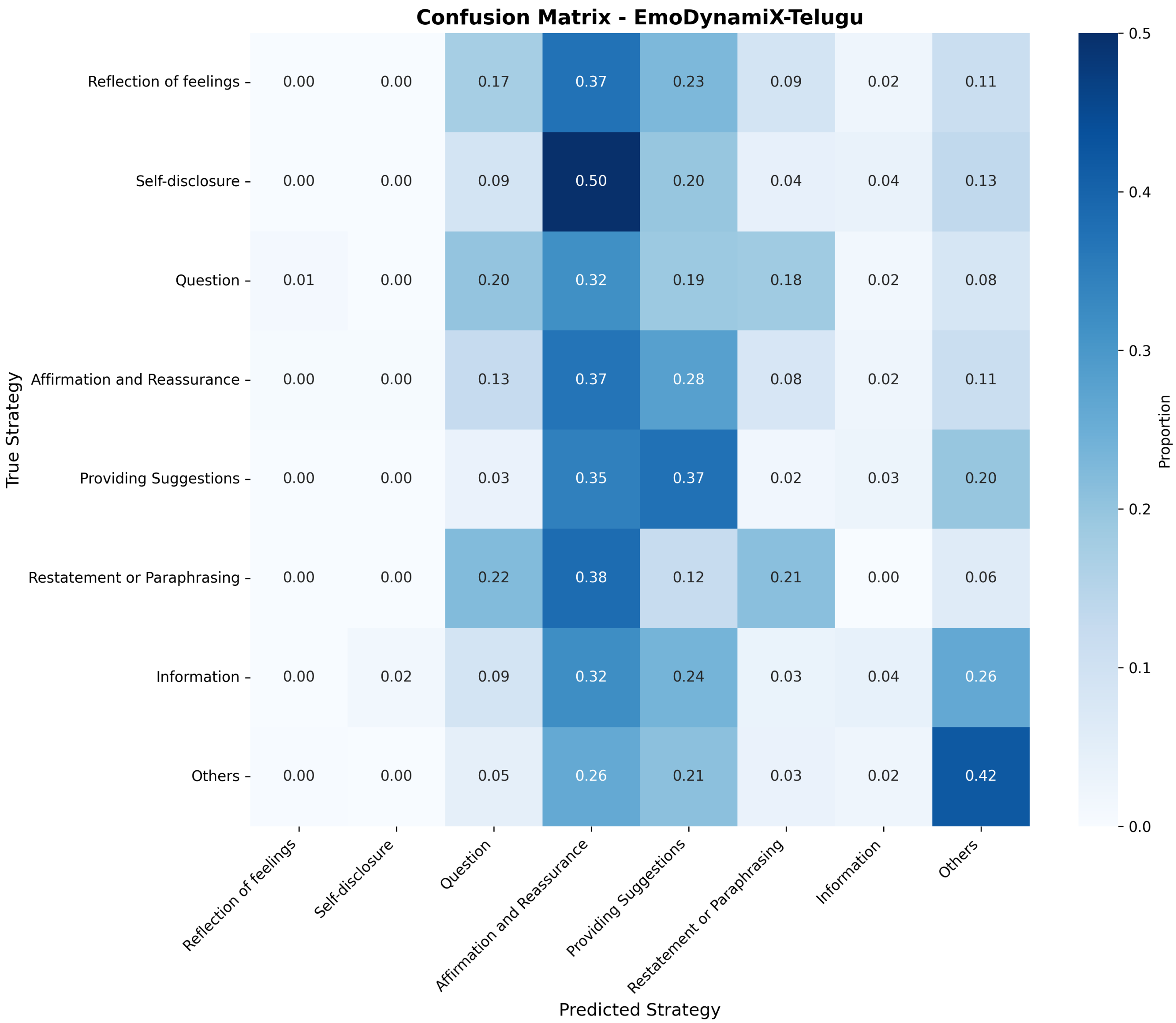

Figure 3: Confusion Matrix for predicted strategies vs true strategies

- **Practical Deployment:** Despite the acceptable F1 scores, our model still makes systematic errors on nuanced emotional strategies. It should be used as a supportive, not autonomous, tool in emotionally sensitive applications.

7.5 Ethical Considerations

Consistent with the goals of the original emoDynamiX work, our system is designed as a *decision support* component rather than a replacement for human counselors. Since the model is trained on translated dialogues and leverages English centric pretrained modules, it may inherit behavioural biases or produce responses resembling human self-disclosure. Care must be taken to prevent misuse and ensure that the system communicates its non-human nature clearly.

All datasets used are anonymized and publicly available. We share our training settings and evaluation code to promote transparency and reproducibility in low-resource socio-emotional modelling.

References

[1] A. et al. Abulimiti. 2023. Large Language Models in Peer Tutoring: Strategy Bias Analysis. In *ACL Findings*.
[2] S. et al. Afantenos. 2015. Graph-based Discourse Parsing on STAC. In *EMNLP*.
[3] N. Asher and A. Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.
[4] A. et al. Braniecka. 2014. Mixed Emotions and Human Affect. *Journal of Happiness Studies* (2014).

[5] J. et al. Chen. 2023. Multimodal Emotional Dialogue GNN. In *EMNLP*.
[6] Y. et al. Chen. 2021. DialogSum: Dialogue Summarization with Discourse Cues. In *ACL*.
[7] M. et al. Chhun. 2024. Explainable Emotional Support with LLMs. In *AAAI*.
[8] Chloé Clavel. 2022. Modular approaches to empathic dialogue. In *COLING*.
[9] Alexis et al. Conneau. 2020. Unsupervised Cross-lingual Representation Learning at Scale. In *ACL*.
[10] Z. et al. Feng. 2021. Modeling Rhetorical Relations for Dialogue Summarization. In *EMNLP*.
[11] T. et al. Fu. 2023. Graph Neural ER. In *EMNLP*.
[12] D. et al. Ghosal. 2019. DialogueGCN. In *EMNLP*.
[13] Xin Hu, Zheng Wei, and Yang Song. 2021. MMGCN: Multimodal Graph Convolution Network for Emotion Recognition. In *ACMMM*.
[14] T. et al. Ishiwatari. 2020. Relation-aware Graph Attention for Dialogue. In *ACL*.
[15] Janani Joshi, Jing Li, and Manuel Pereira. 2021. EMPDIALOGUE: A Benchmark for Multi-turn Empathetic Dialogue Generation. In *SIGDIAL*.
[16] M. et al. Kang. 2024. Strategic Planning Improves Emotional Support Dialogue Generation. In *NAACL*.
[17] S. et al. Kim. 2021. Emotion-BERT for Conversational Emotion Recognition. In *EMNLP*.
[18] Y. et al. Li. 2023. DialGraph: Discourse-Aware Dialogue Emotion Recognition. In *ACL*.
[19] Hanxiao Liu, Beidi Chen, Tianyu Zhao, and Quanquan Gu. 2022. Graph Neural Networks with Learnable Structural Representations. In *NeurIPS*.
[20] Siyan Liu, Tao Chen, Xiang Ren, and Zhou Yu. 2021. Towards emotional support dialog systems. In *ACL*.
[21] X. et al. Lu. 2024. Transparent AI for Counseling Dialogues. *TACL* (2024).
[22] Q. et al. Ludan. 2023. Interpreting Black-box LLMs. In *ICLR Workshops*.
[23] W. Mann and S. Thompson. 1988. Rhetorical Structure Theory. *Text* (1988).
[24] J. et al. Perret. 2016. Integer Linear Programming for Discourse Parsing. In *ACL*.
[25] S. et al. Poria. 2019. MELD: A Multimodal Multi-Party Emotional Dialogue Dataset. In *ACL*.
[26] Vandit Raut, Akshata Patil, and Manish Shrivastava. 2023. SUPPORTER: Modeling Support Strategies in Emotional Support Conversation. In *EMNLP*.
[27] Lisa et al. Shen. 2021. Directed Acyclic Emotion Graphs for ERC. In *ACL*.
[28] W. et al. Shi. 2019. Deep Sequential Discourse Parsing. In *ACL*.
[29] S. et al. Wang. 2020. Conversational Graph Modeling. In *ACL*.
[30] Wenyue et al. Wu. 2022. Pushing on Textual and Graph Reasoning for Emotional Support Conversations. In *ACL Findings*.
[31] Y. et al. Yang. 2023. Commonsense-Aware Dialogue Graphs. In *ACL*.
[32] S. Zahiri and J. Choi. 2018. Emotion Detection in Conversations: EmoryNLP Dataset. In *NAACL*.
[33] K. et al. Zhang. 2023. Dialogue Graph Networks for Emotion Modeling. In *EMNLP*.
[34] S. et al. Zheng. 2023. Benchmarking LLMs for Emotional Support Conversations. In *EMNLP*.
[35] R. Zwaan. 2012. System architecture for emotional support chatbot. In *KEOD*.
[36] R. Zwann. 2012. Microworld simulations for emotional dialogue systems. In *IVA*.

A Datasets

A.1 ESC Datasets

Following are the two public ESC datasets we use to evaluate our method.

ESConv. (Liu et al., 2021) utilizes eight dialogue strategies: Reflection of Feelings, Self Disclosure, Question, Affirmation and Reassurance, Providing Suggestions, Restatement or Paraphrasing, Information, and Others. The distribution of these strategies is depicted in Figure 5. ESConv collects user feedback scores after every few speaker turns to evaluate the effectiveness of emotional support. Notably 79.9% of the scores are above 4 (Good), indicating a high overall quality of the conversations. To ensure fair comparison with prior baselines, we perform no filtering. The top-3 topics include *Ongoing depression*, *Job crisis*, and *Break-up with partner*.

AnnoMI. (Wu et al., 2022) categorizes therapist behaviors into four high-level types: Reflection, Question, Input, and Other. These are further decomposed into nine fine-grained strategies: Simple Reflection, Complex Reflection, Open Question, Closed Question, Information, Advice, Giving Options, Negotiation/Goal-setting, and Other. Due to data sparsity, we merge *Advice*, *Giving Options*, and *Goal-setting* into a single strategy: *Provide Suggestion*. The distribution is in Figure 5. AnnoMI contains 110 high-quality (82.7%) and 23 low-quality (17.3%) dialogues. We exclude low-quality ones for training.

A.2 Datasets for Pre training Expert Models

STAC. (Asher et al., 2016) is used to pre-train our discourse parser. It contains 1,081 multi-party game dialogues with 16 discourse dependency types such as Comment, Clarification Question, Elaboration, Explanation, Contrast, and more.

DailyDialog. (Li et al., 2017) is used to pre-train the ERC module. It contains 13,118 multi-turn dialogues covering everyday scenarios. Emotions include Ekman's six basic categories plus Neutral.

B Definitions of ESC Strategies

B.1 Strategies in ESConv

- **Question**: Asking for information to help the seeker articulate their issues.
- **Restatement or Paraphrasing**: A concise rephrasing of the seeker's statements.
- **Reflection of Feelings**: Describing the help-seeker's feelings to show empathy.
- **Self-disclosure**: Sharing similar experiences or emotions.
- **Affirmation and Reassurance**: Affirming ideas and strengths to encourage the seeker.
- **Providing Suggestions**: Offering ways to overcome difficulties.
- **Information**: Providing useful factual information or resources.
- **Others**: Any remaining strategies.

B.2 Strategies in AnnoMI

- **Question open**: Encourages elaboration and self-exploration.
- **Question closed**: Gathers specific, concise information.
- **Reflection simple**: Communicates basic understanding.
- **Reflection complex**: Shows deeper insight into the user's perspective.
- **Provide suggestion**: Combined category of Advice, Options, Goal-setting.
- **Provide information**: Supplies factual or helpful details.
- **Other**: Other conversational behaviors.

C Baselines

ChatGPT. (OpenAI, 2023) is a 175B parameter model trained with RLHF. Prompting follows Figure 6, using strategy definitions from Appendix B.1/B.2. Two-shot examples are retrieved via SBERT [?] similarity.

LLaMA3 (70B & 8B). (Meta, 2024) are instruction-tuned models optimized for dialogue. The 70B variant uses the same prompting templates; the 8B variant uses flattened inputs and LoRA [?] for fine-tuning.

RoBERTa. (Liu et al., 2019) is an improved BERT model using dynamic masking and larger training batches.

BART. (Lewis et al., 2020) is an encoder decoder model trained via text reconstruction.

MISC. (Tu et al., 2022) uses BlenderBot with COMET commonsense knowledge.

MultiESC. (Cheng et al., 2022) is a BART-based ESC system with lookahead search.

KEMI. (Deng et al., 2023) integrates HEAL case knowledge through COMET.

TransESC. (Zhao et al., 2023) uses BlenderBot with emotion modeling and COMET-based state transitions.

D Implementation Details

D.1 Preference Bias Score

Given confusion matrix w_{ij}, the iterative preference update is:

$$p_i' = \frac{\sum_j (w_{ij} p_j)/(p_i + p_j)}{\sum_j w_{ji}/(p_i + p_j)}$$

All preferences initialize as $p_i = 1$. We run 20 iterations. Preference Bias is:

$$B = \sqrt{\frac{\sum_{i=1}^{N} (p_i - p)^2}{N}}$$

D.2 Submodule Details

Discourse Parser. Trained with the same setup as Chi and Rudnicky (2022): LR = 2e-5, batch size = 4, 4 epochs, 10% warm-up. Achieves 59.0 F1.

ERC Module. Pre-trained on DailyDialog with LR = 2e-5, 500 warm-up steps, weight decay = 1e-3, trained 12k steps. Achieves accuracy 82.26, macro F1 53.0, weighted F1 83.54.

D.3 Hyperparameters

For EmoDynamiX-Telugu (`XLMRHeterogeneousGraphTelugu`), we did train the full model end-to-end on top of frozen XLM roberta Large and a frozen Telugu ERC encoder, including the mixed emotion module, the S/E embedding spaces, the RGAT based heterogeneous graph layers, and the MLP classification head. We use a batch size of 8 and a learning rate of 3e−5 with AdamW optimizer, a linear warm-up over the first 300 steps, and a weight decay of 1e−2. The heterogeneous graph hidden dimension is set to 256 (i.e., `hg_dim` = 256), and the initial mixed-emotion temperature is $\tau_0 = 0.1$ (from the `erc_temperature` argument, rescaled internally). We employ 3 RGAT layers and keep both the XLM-R encoder and the Telugu ERC encoder frozen throughout training.

EmoDynamiX-Telugu is trained for 1500 steps on ESConv Telugu on a single NVIDIA RTX 4090 GPU, using mixed precision and, when applicable, precomputed ERC logits and discourse parses for efficiency.

E Supplementary Analysis

E.1 Impact of the Initialization of τ

Figure 7 shows the effect of the initial temperature τ on mixed emotion distributions. Lower τ sharpens distributions and yields best results. Very small τ approximates one-hot vectors; extremely large τ softens the distribution excessively and hurts performance.

E.2 Output Statistics of the ERC Module

Table 3 shows the ERC output distribution on ESConv vs. original DailyDialog. ESConv has significantly higher emotional intensity.

Emotion	ESConv	DailyDialog
Anger	1.83	0.99
Disgust	0.70	0.34
Fear	0.61	0.17
Joy	20.17	12.51
Sadness	17.17	1.12
Surprise	0.31	1.77
Neutral	59.22	83.10

Table 3: Comparison between ERC output on ESConv and label distribution of DailyDialog.

E.3 Confusion Matrix of EmoDynamiX on ESConv

(Include your confusion matrix figure here.)