NAACL 2025

# EmoDynamiX: Emotional Support Dialogue Strategy Prediction by Modelling MiXed Emotions and Discourse Dynamics

Chenwei Wan, Matthieu Labeau, Chloé Clavel

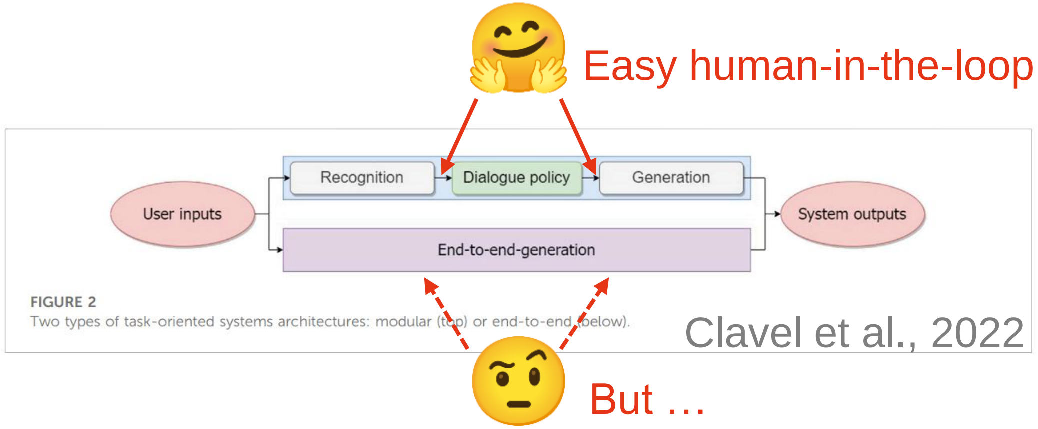
TELECOM Paris
INSTITUT POLYTECHNIQUE DE PARIS
Inria

# To Think about Conversational Agents

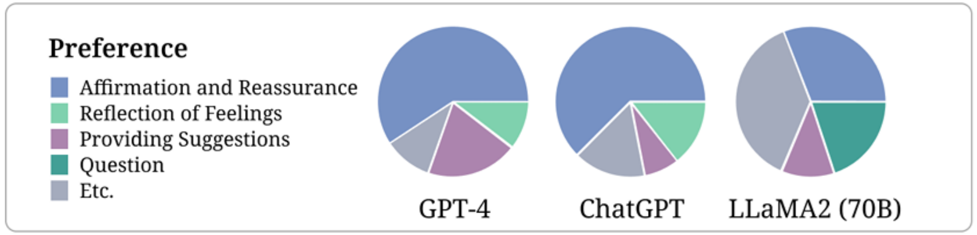
Consider two patterns for conversational agents: modular vs end-to-end.

LLMs make end-to-end an easy option, but they lack **transparency** and **controllability**.

Besides, LLMs tend to overly rely on certain dialog strategies.




Clavel et al., 2022




LLMs' heavy preference bias over dialog strategies (Kang et al., 2024)

# Our Solution

- Decouple strategy prediction from generation: a modular point of view
- We present EmoDynamiX: a dedicated dialog strategy prediction framework using explicit cognitive modeling

# Social Intelligence Matters

The ability to perceive human mental states and choose the appropriate social behavior is essential.

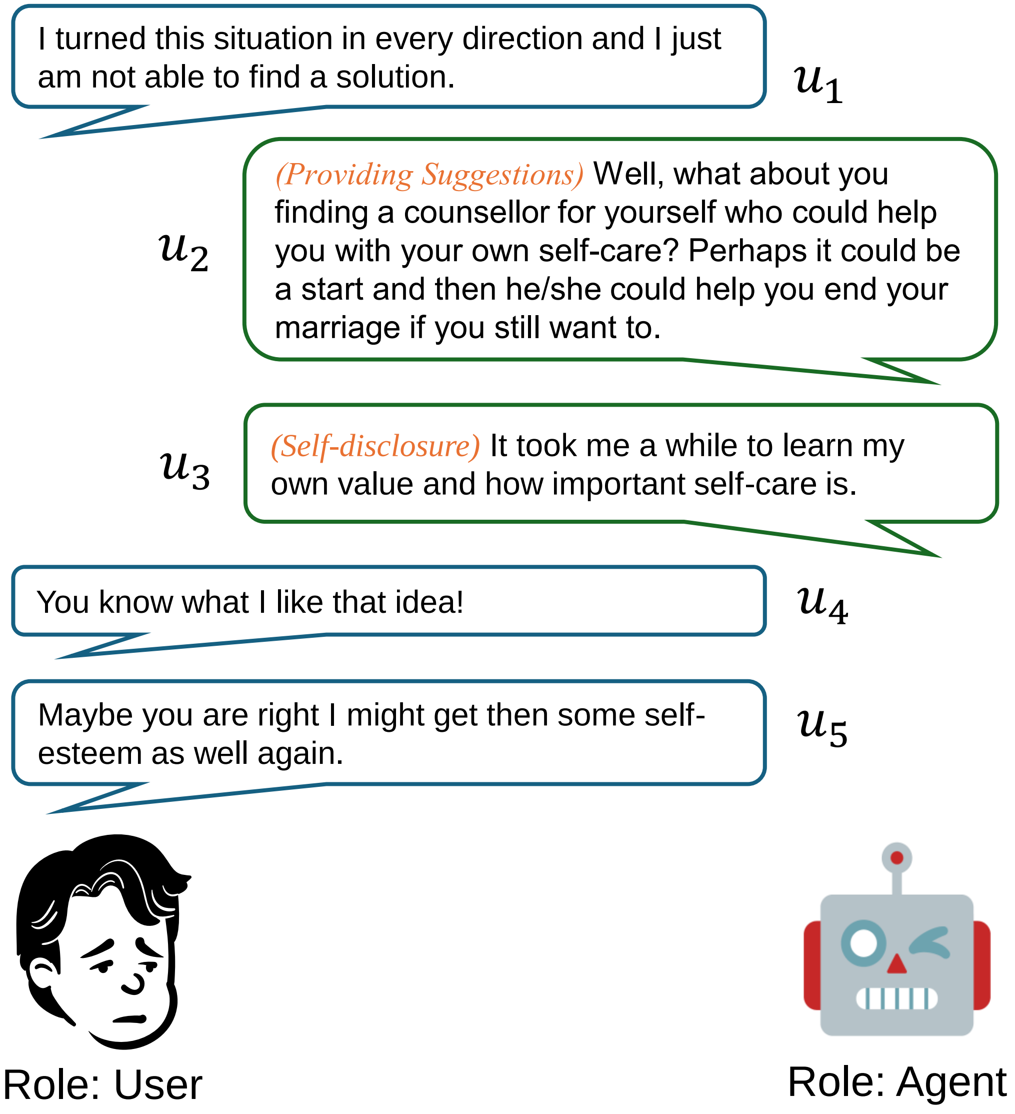
However, LLMs are inherently task-oriented (Abulimiti et al., 2023) and significantly lag behind humans in modeling social cognitive processes (Chen et al., 2024).

Created with DALL·E 3

# Task

**Input**: Dialog History

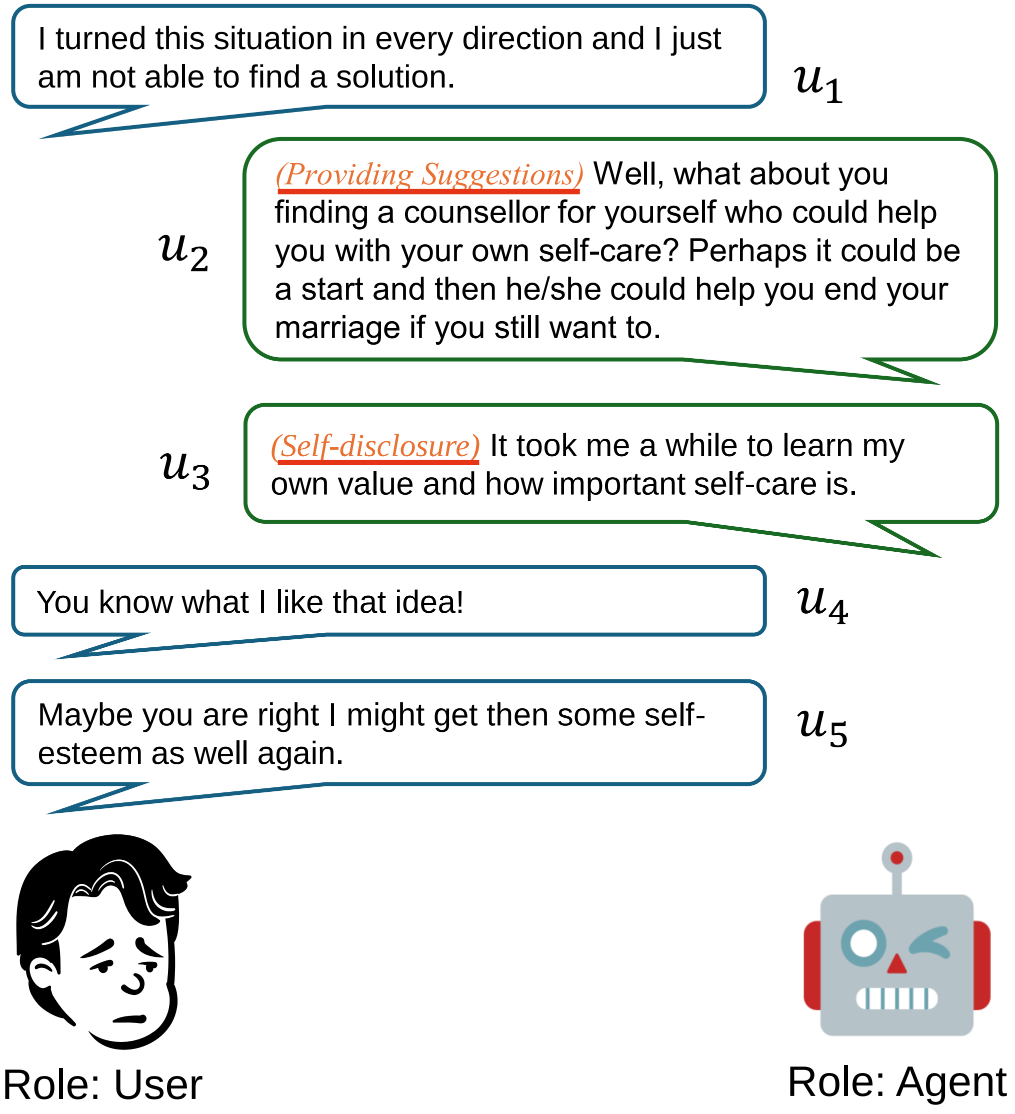
- Context window size = 5



# Task

**Input**: Dialog History

- Context window size = 5
- The past strategies applied by the agent are considered

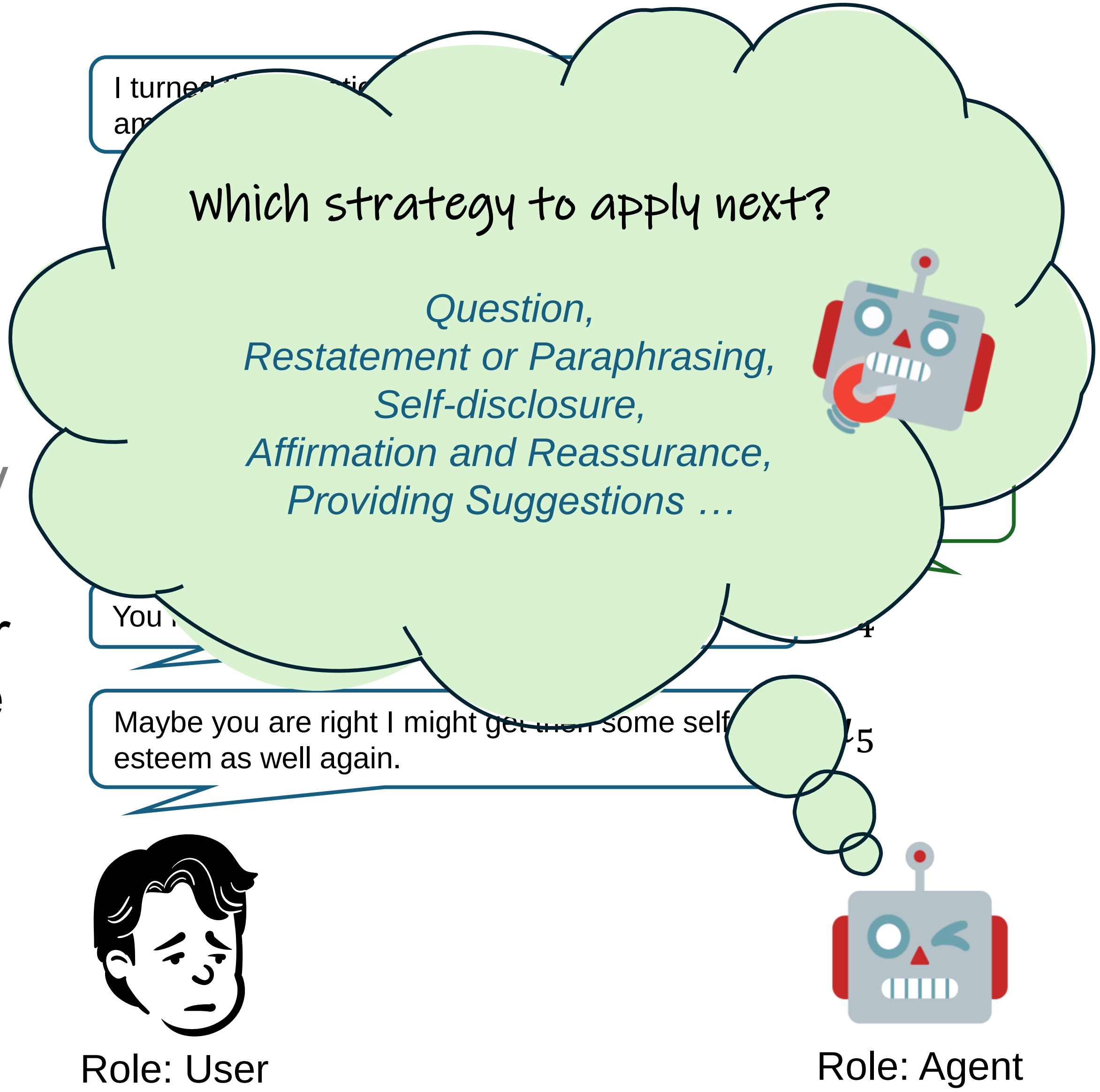


# Task

**Input**: Dialog History

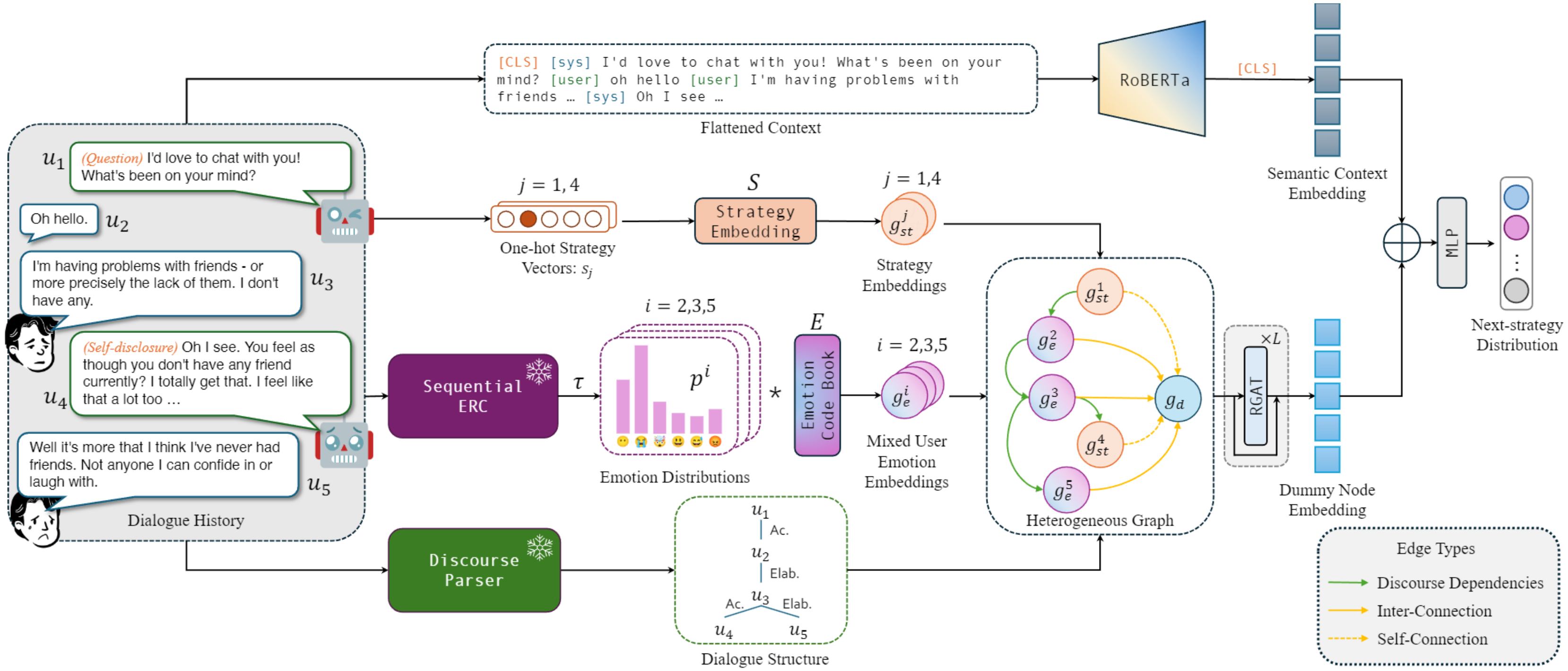
- Context window size = 5
- The past strategies applied by the agent are considered

**Output**: Next dialog strategy for controlling the generation of the agent's next response



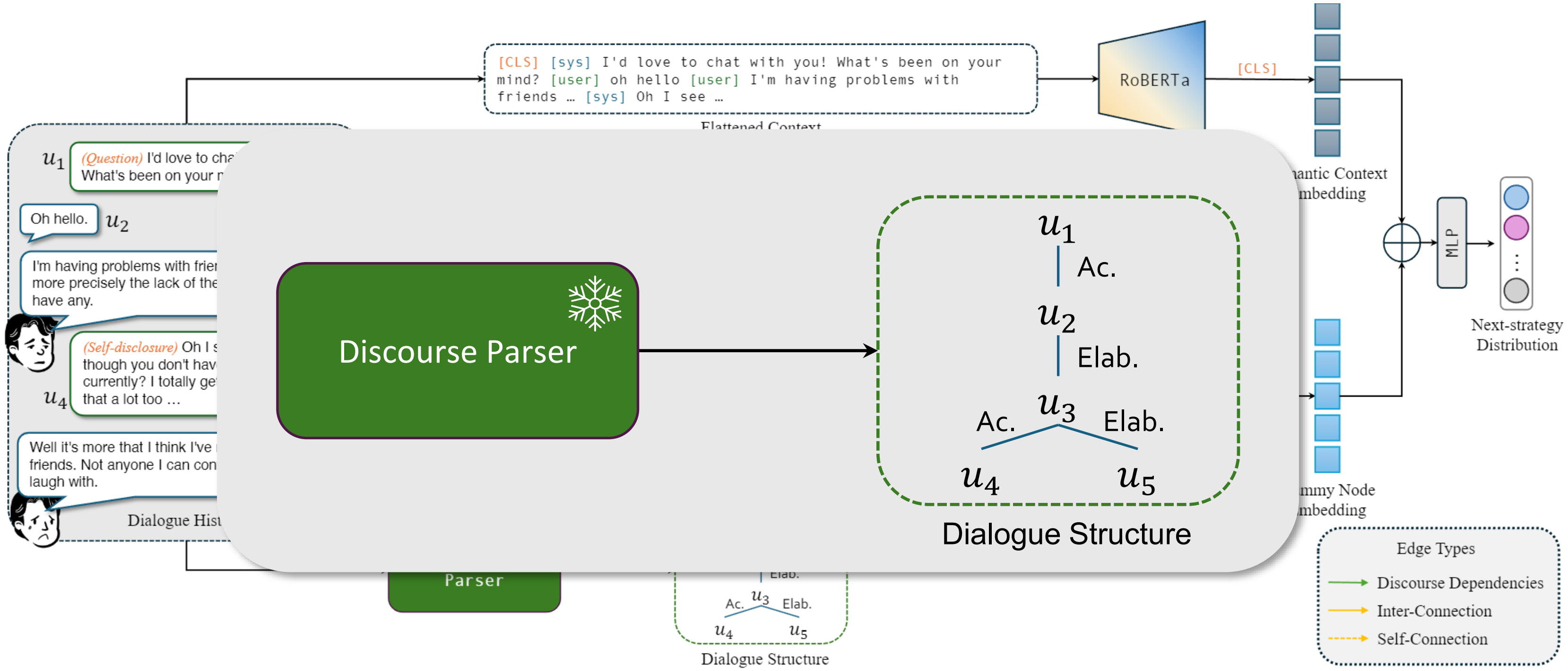
# Methodology

## Overview



# Methodology

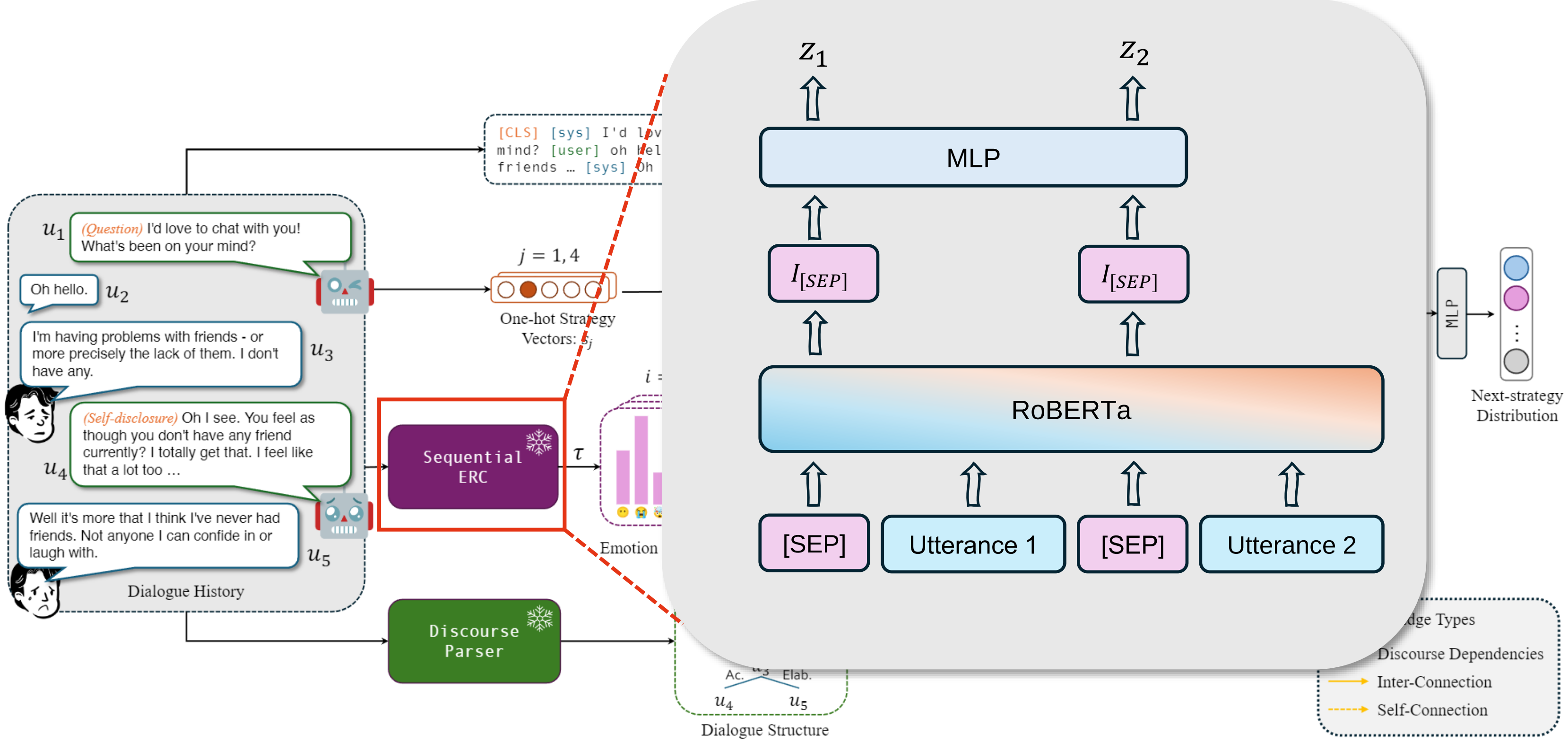
Pre-trained discourse parser for mining the dependencies between speaker turns, serving as the backbone of our explicit discourse dynamics model




Example dependency types: *Acknowledgment, Elaboration, Correction …*

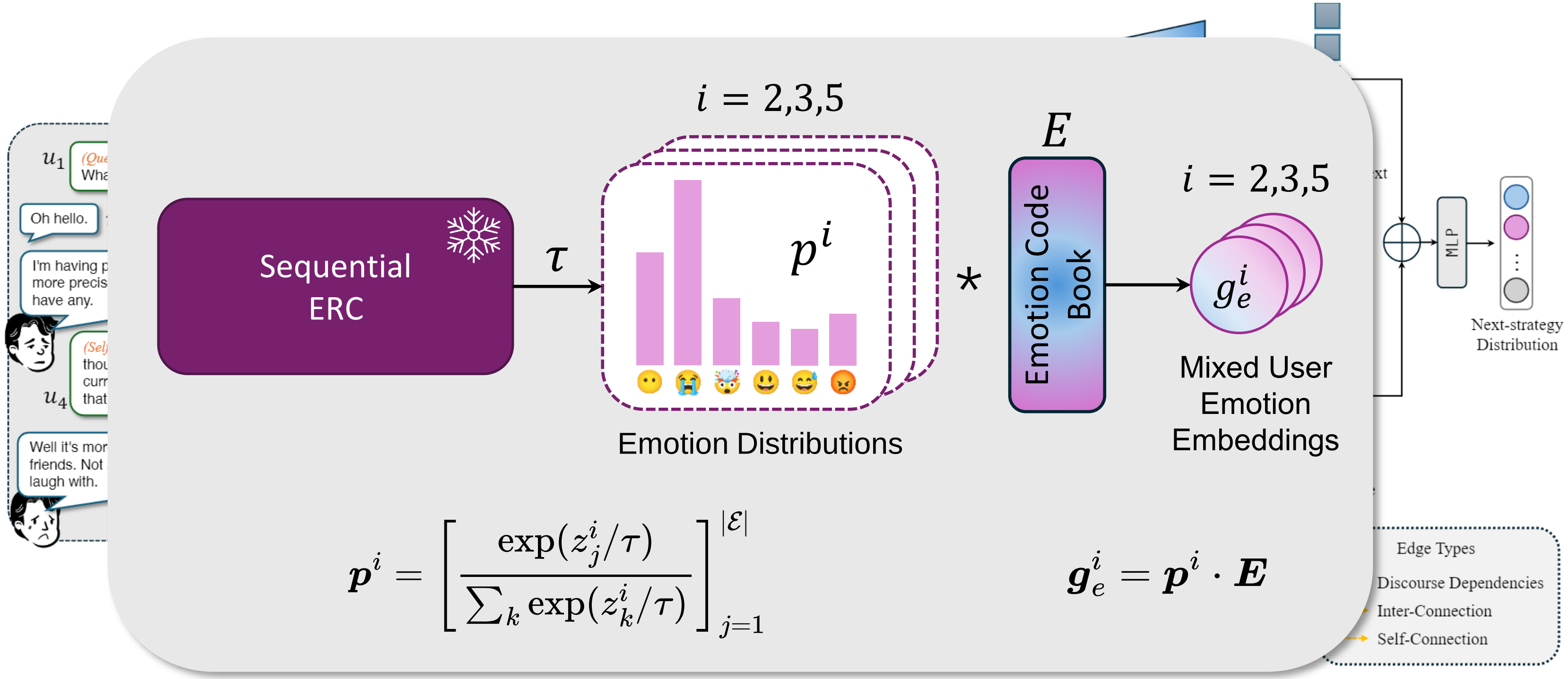
# Methodology

Pre-trained emotion recognition model for mental state inference



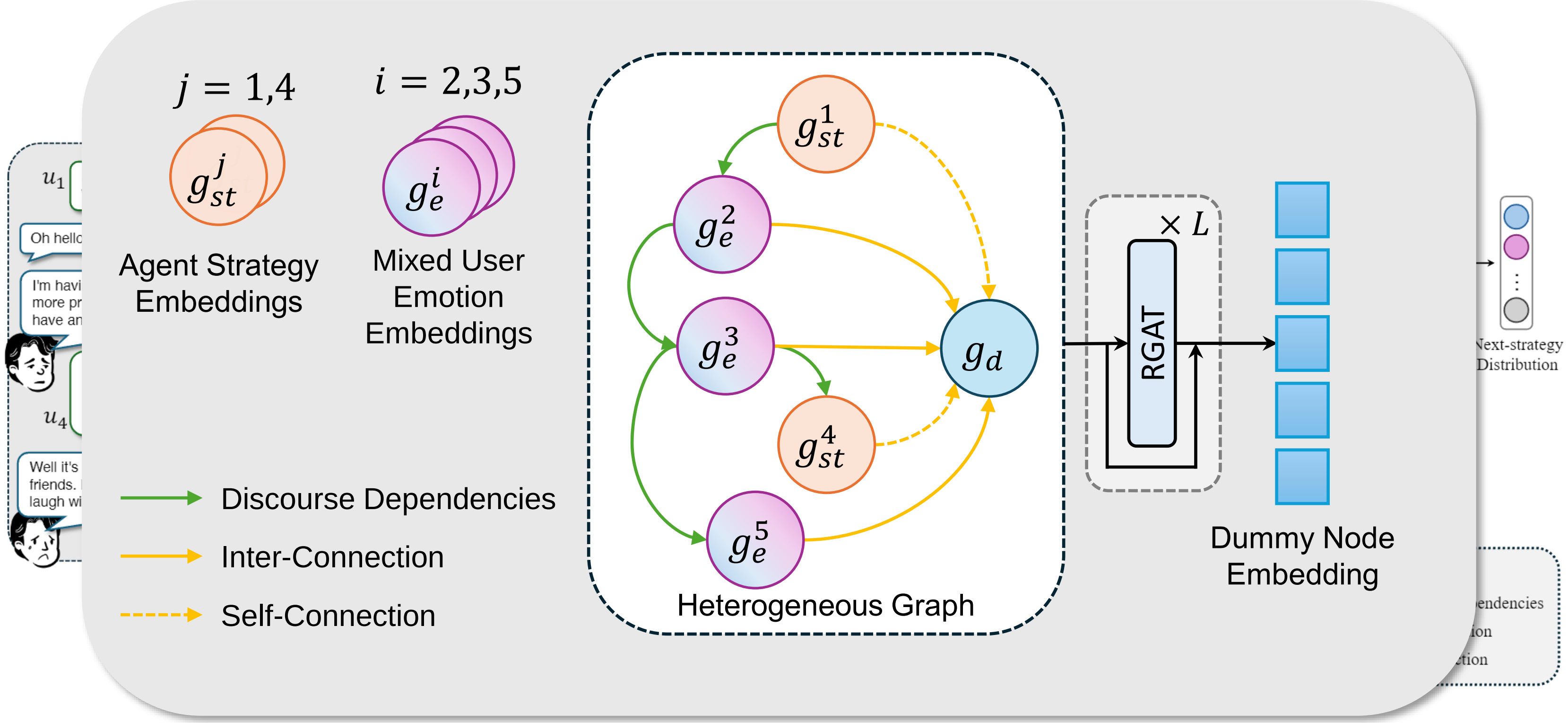
# Methodology

Fine-grained mental states are represented with mixed emotions: we use temperature-controlled *emotion distributions* instead of *discrete labels*



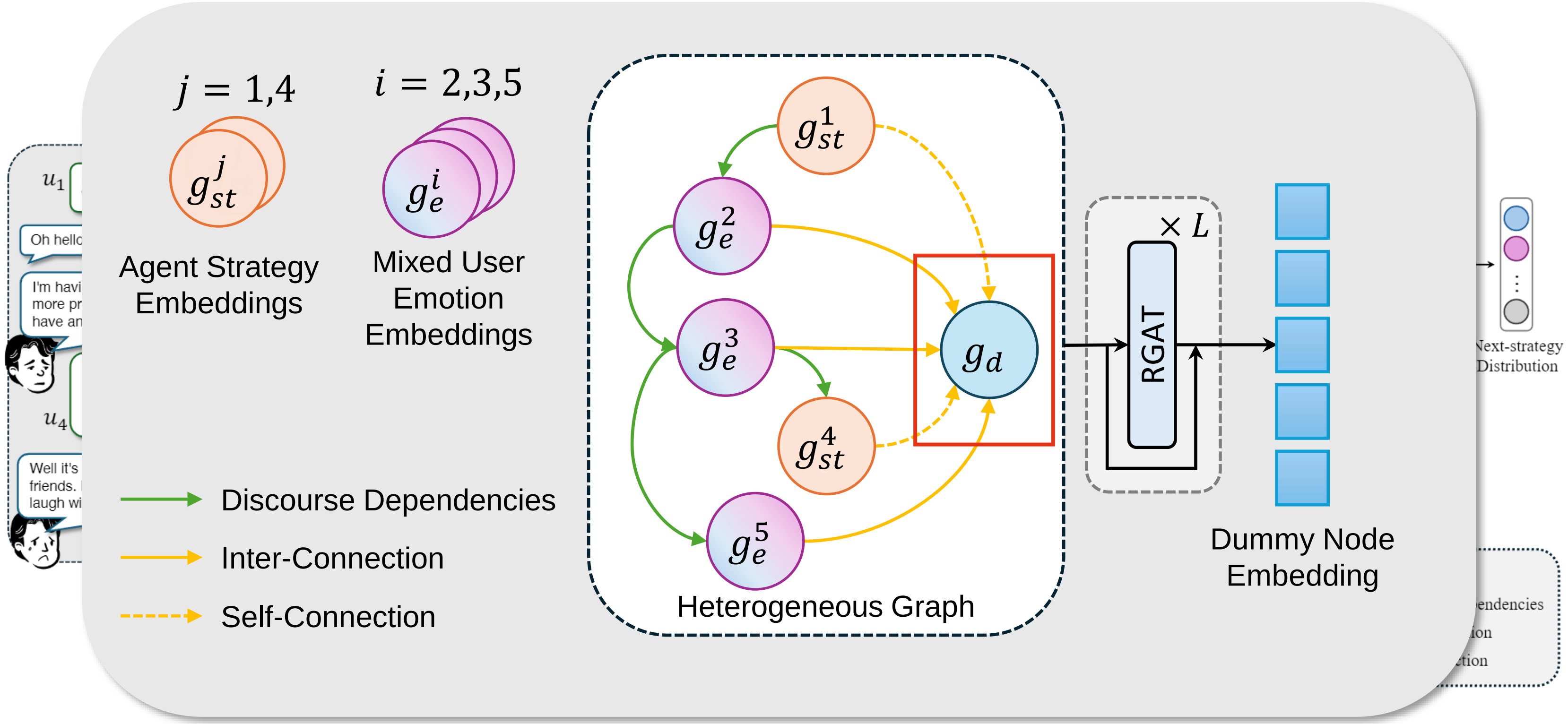
# Methodology

Graph neural network for representing the speaker-aware conversational dynamics



# Methodology

## Dummy node for information aggregation



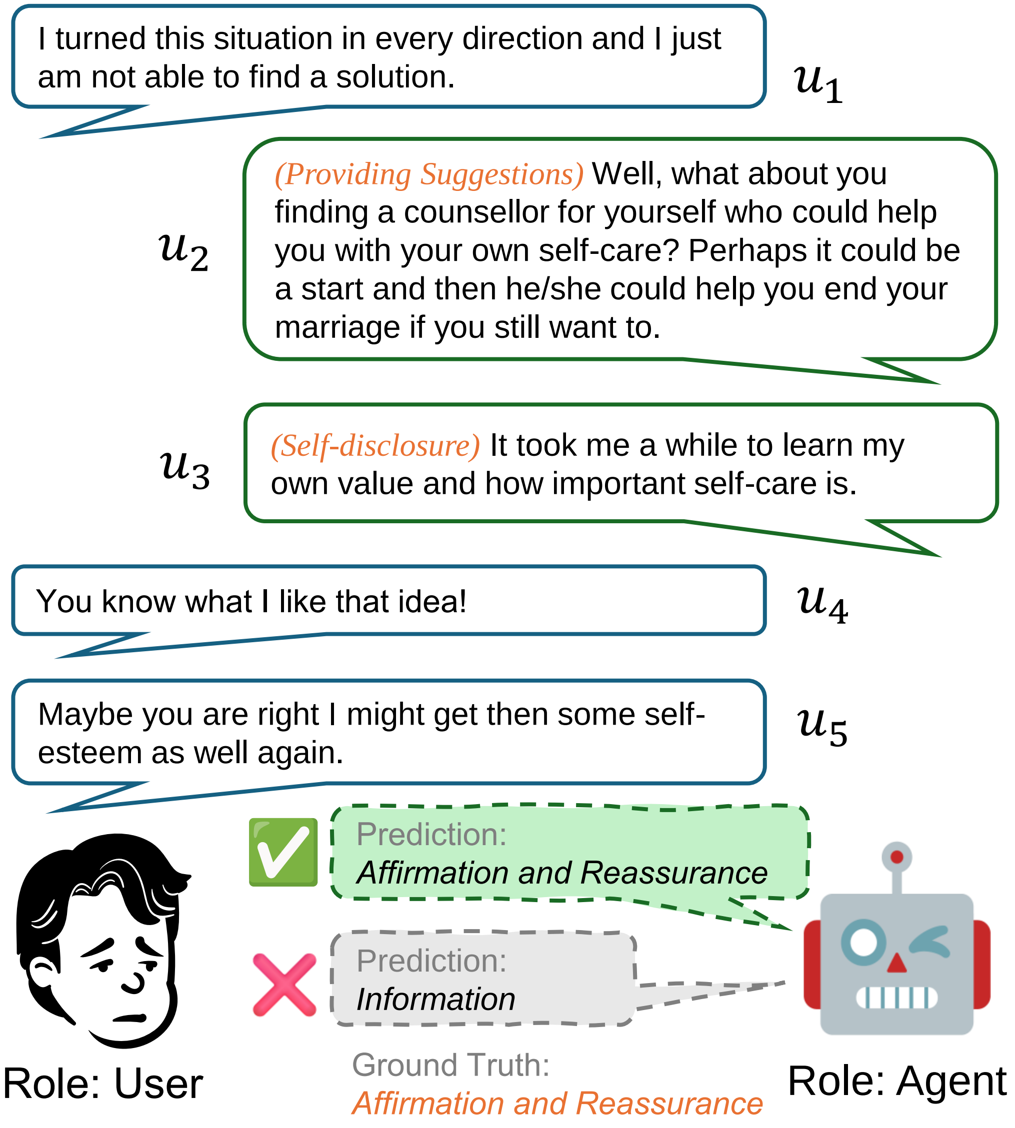
# Experimental Setup

ESC Datasets: **ESConv** (Liu et al., 2021) and **Anno-MI** (Wu et al., 2022)

- Label: one GT strategy per dialog snippet

Evaluation metrics:

- Proficiency: Weighted & Macro F1
- Preference Bias score: how unevenly the model favours certain strategies over others (Kang et al., 2024)



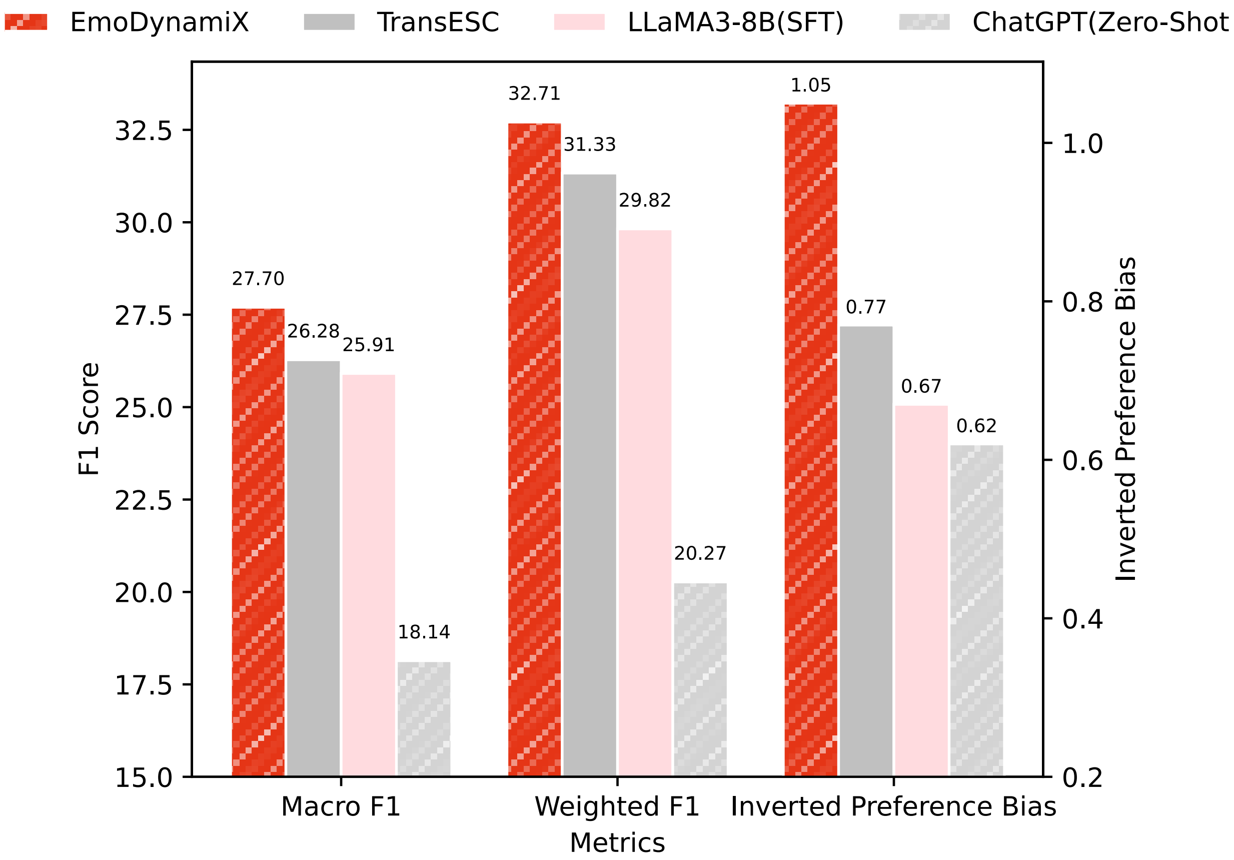
# Main Results

EmoDynamiX outperforms:

- **SOTA ESC frameworks**: TransESC (Zhao et al., 2023), etc.
- **Finetuning SOTA LLMs**: LLaMA3-8B (Meta, 2024), etc.
- **Prompting SOTA LLMs**: ChatGPT (OpenAI, 2023), etc.

Across all metrics on ESConv and AnnoMI.

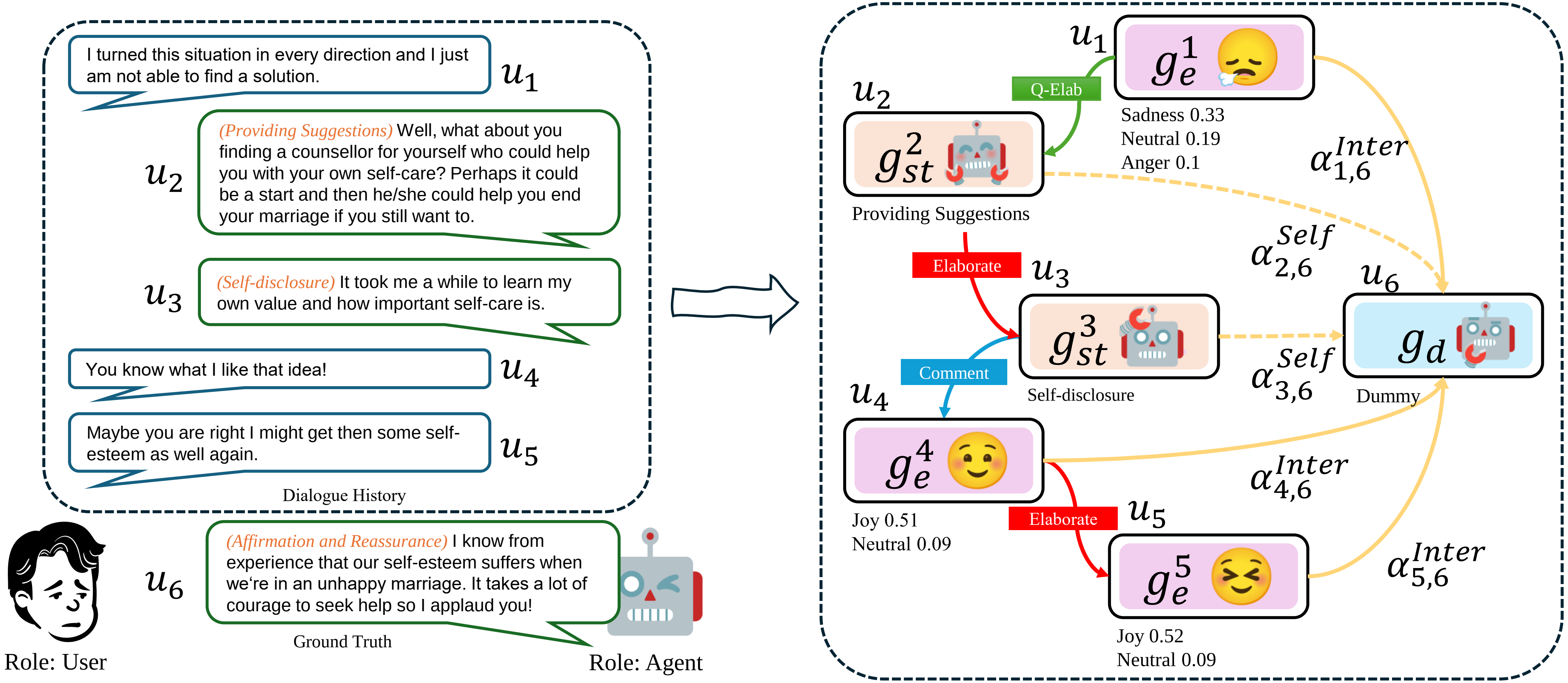
## Results on ESConv




Inverted Preference Bias = 1.5 – Preference Bias

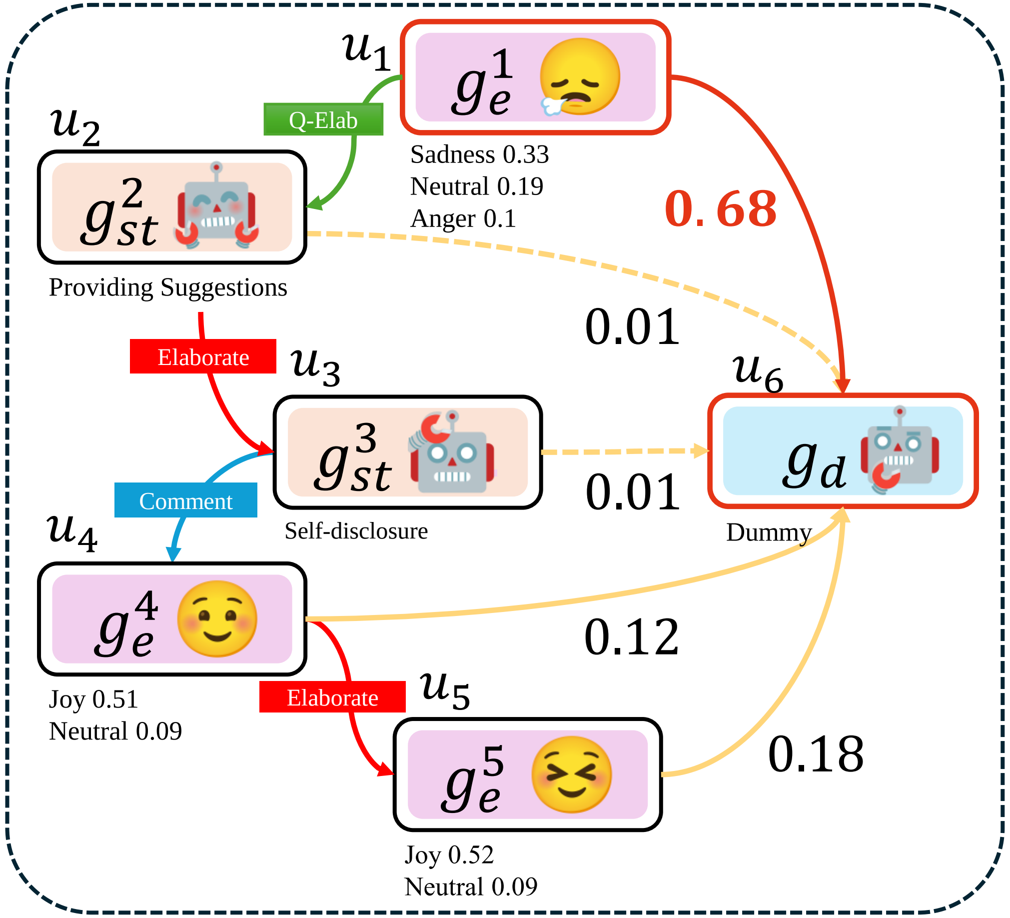
# Case Study

## Explicitly model the discourse dynamics with a heterogeneous graph



# Case Study

## Extra transparency are provided with back-tracing attention weights

# Case Study

Mixed-emotion module can model a large set of subtle emotional expressions by combining primary emotions

$u_1$ I turned this situation in every direction and I just am not able to find a solution.

☝️ Emotion here can be inferred as ***Frustration,*** which does not belong to common emotion categories in ERC datasets.

$g_e^1$ 😮‍💨

Sadness 0.33
Neutral 0.19
Anger 0.1

☝️ Primary emotion categories from DailyDialog (Li et al., 2023)

# Conclusion

- Focusing on the policy layer of modular dialog agents is a necessary task due to the need of controllability.
- Explicit cognitive modeling can significantly improve proficiency of decision-making in human-AI conversation, while providing extra transparency.
- Our mixed-emotion module helps to model fine-grained mental states, which can better support decision-making than general commonsense knowledge.

# Experimental Setup

Preference Bias score:

- *Preference* $p_i$ indicates the degree to which the model favours strategy $i$ over others. It is calculated iteratively using the confusion matrix:

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

- *Preference Bias* is the standard deviation of $p$:

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