

# Sentiment Aided Graph Attentive Contextualization for Task Oriented Negotiation Dialogue Generation

Aritra Raut<sup>1</sup>, Sriparna Saha<sup>2</sup>, Anutosh Maitra<sup>3</sup> and Roshni Ramnani<sup>3</sup>

<sup>1</sup>Ramakrishna Mission Vivekananda Educational and Research Institute, India

<sup>2</sup>Department of Computer Science and Engineering, Indian Institute of Technology Patna, India

<sup>3</sup>Accenture, India

## Abstract

Over the past several years, the demand and popularity of using virtual assistants to finish jobs like service scheduling and online shopping have increased. While keeping the user's request in mind, an effective task-oriented virtual agent must strive to improve the seller's profit. Therefore, in order to achieve the best possible trade-off between the parties, this form of virtual agent has to have strong negotiating abilities. Although current conversational agents are quite good at making fluent sentences, they are still unable to use strategic thinking. In order to more effectively contextualize the choice of the next set of negotiation methods while producing answers, we develop Nego-GAT, an end-to-end negotiation system that includes sentiment information and graph attention embedding into GPT-2. Our self-supervised model outperforms earlier cutting-edge negotiation models in terms of both the precision of strategy/dialogue act prediction and the caliber of the generated dialogue responses<sup>1</sup>.

## 1 Introduction

Recent natural language processing research has centered on the creation of models for conversational agents, which have a variety of applications in the business, sales, and healthcare sectors. Depending on their intended purpose, conversational agents can be divided into two categories: chitchat agents and task- or goal-oriented virtual agents. Chit-chat agents interact with users as a friend to meet their need for companionship and for conversation, while the former aims to help users complete tasks.

Conversational bots for tasks have recently become more common in the field of natural language generation (NLG). Customers may get assistance from these agents with a variety of tasks,

<sup>1</sup>all codes are available at <https://github.com/aritraraut/Nego-GAT/tree/master>

including booking hotel rooms, buying tickets, and more. With the use of implicit encoder-decoder designs (Sordoni et al., 2015; Li et al., 2016) or explicit semantic information, such as slot-value pairs (Young, 2006; Larionov et al., 2018), modern conversation systems have made significant progress in characterizing the history and structure of discussion. The user's need may not be met, or there may be other barriers, such as financial constraints, that prevent the user from completing a purchase, which may prevent these agents from being successful. In this case, one of the best strategies for assisting both parties in reaching an understanding is bargaining (An example has been shown in Fig 1).

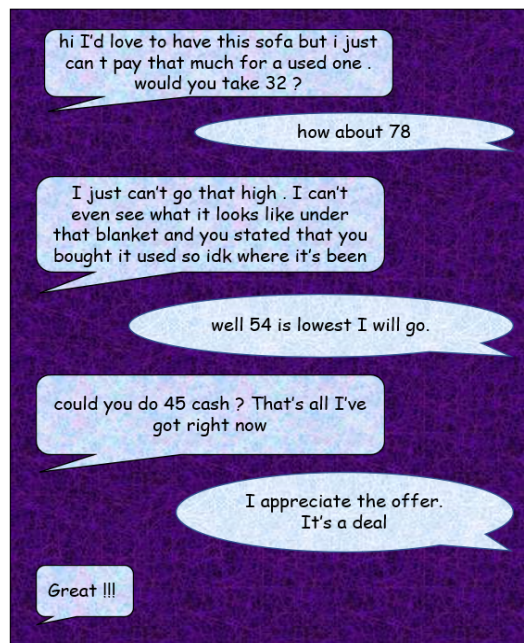


Figure 1: An example of negotiation conversation

Users openly state their intentions in these tasks, which enables computers to map the utterances to specific intent slots (Li et al., 2020). However, such mapping is less obvious in difficult non-collaborative tasks like negotiation (He et al.,

2018) and persuasion (Wang et al., 2019) since the user’s intent and the most effective strategies are concealed. Selecting the appropriate dialogue act and negotiating approach is essential for generating appropriate responses. The majority of earlier research on negotiating dialogues really concentrated on improving conversation techniques, ranging from high-level task-specific strategies (Lewis et al., 2017) to more precise task execution planning (He et al., 2018), to fine-grained planning of language outputs given strategic options (Zhou et al., 2019a). Joshi et al. (2021) proposed a framework for modeling intricate negotiation strategies that makes use of intermediate structures and Graph Attention Networks (GAT) to make the model understandable. Their model makes use of the recently proposed hierarchical graph pooling-based methodology to comprehend the relationships between negotiating techniques, such as conceptual and verbal strategies and conversation activities, and their relative value in selecting the best sequence.

All of these methods frequently employ various modules to precisely simulate different components of the conversation history. The pipeline strategy’s apparent downside is that subsequent sub-tasks could be damaged by error propagation from cascaded components (Liu and Lane, 2018). Thus, developing an end-to-end system is essential in order to not only give concise responses but also to decide on the best approach to negotiations. Yang et al. (2021) suggested UBAR, a fully end-to-end chat system. In a straightforward task-oriented conversation generation setting, this model draws belief states from the context and also independently decides actions and generates responses.

In keeping with the concept put forward by Joshi et al. (2021) and Yang et al. (2021), we present our model Nego-GAT, which integrates GAT (Graph attention) embeddings into the large pre-trained model GPT-2 (Radford et al., 2019) to enhance contextualization while producing negotiation answers. On the other hand, in the case of negotiation dialogues, sentiment plays an essential role while generating responses. For instance, Wang et al. (2019) included user sentiment to make an effective user-adaptive system. With this in mind, we also retrieved the sentiment information from the user text and used it as an additional piece of context-specific data to further enhance our model.

Using the CraigslistBargain (He et al., 2018)

dataset, we have refined our model at the session level, taking into account the user utterance, sentiment, belief state<sup>2</sup>, bargaining strategy, conversation act, and agent reaction for each dialogue turn. On the same dataset, we ran a number of experiments, and we found that our model outperformed all previous models virtually universally. The following list outlines the study’s contributions:

1. This is the first approach towards building an end-to-end task-oriented model that incorporates GAT embeddings into the large pre-trained model like GPT-2 in an effort to enhance negotiation dialogue generation by choosing the appropriate sequence of strategies. We also incorporate sentiment tokens as an extra bit of information into the context of the same goal as mentioned.
2. Using a standard dataset on negotiation exchanges (CraigListBargain He et al. (2018)), we assess the proposed model. The outcomes demonstrate that our model routinely performs better than the most recent models.

## 2 Related Works

Our main goal is to create an end-to-end system that can extract belief states, select the appropriate approach, and produce a strong negotiating response. So, we’ve highlighted a few related and relevant works in this section.

**Task oriented virtual agents:** A number of sequence-to-sequence based dialogue (Li et al., 2017) generation techniques have been proposed, which use RNN (Recurrent Neural Network) units (LSTM/GRU) to encode dialogue context and construct replies using the encoded data. After that, the use of pre-trained models, like GPT-2 (Radford et al., 2019), has become more widespread in recent years. Budzianowski and Vulić (2019) first emphasized the ability to fine-tune all relevant information in plain text on GPT-2 in the context of task-oriented interaction. Later, (Yang et al., 2021) were able to develop an end-to-end task-oriented agent that outperformed all earlier models with just a little tinkering. Yet, in a few recent methods, researchers (Chiu et al., 2022; Sun et al., 2020) have tried to bridge the gap between chit-chat and task-oriented dialogue agents in an effort to make the

---

<sup>2</sup>We have performed belief state annotation on the CraigListBargain dataset (He et al., 2018)

task-oriented conversation more fascinating and appealing.

**Persuasive dialogue generation :** On the other hand, attempts were made to add persuasion (which is basically the superset of negotiation) into the NLG(Natural Language generation) module. According to Petty and Cacioppo’s Elaboration Likelihood Model (ELM) (Petty and Cacioppo, 1986), a person’s persuasion is dependent on varying degrees of processing information and a persuasive environment. The Persuasion Knowledge Model (PKM) proposed by Friestad and Wright (1994) proposes that scientific and common persuasion knowledge are interconnected. The paper (Qiu and Zhang, 2021) presents a personalized end-to-end task-oriented conversation system that uses a memory network to provide appealing and persona-consistent responses. In recent articles, the researchers (Tiwari et al., 2021, 2022) highlighted the DST(Dialogue State Tracker) module’s application in task-oriented conversation agents to carry out persuasion in order to effectively identify and respond to changing user demands. Other than this, a meta-learned end-to-end model using GPT-2 has been proposed (Raut et al., 2022, 2023) to produce persuasive dialogues in task oriented dialogue generation setting.

**Negotiation dialogue generation:** There hasn’t been much literature developed in this area. He et al. (2018) suggested a scenario for negotiation that can take advantage of semantic and tactical events. To learn conversation structure, Zhou et al. (2019b) employed unsupervised learned FSTs (Finite State Transducer). Although this method clearly incorporates pragmatic tactics, it does not take advantage of the expressive power of neural networks. In contrast, Joshi et al. (2021) tried to capture the interplay between the negotiation strategies in successive turns using Graph Attention Networks(GAT), and finally used a different decoder module to generate responses. In addition to the possibility of error propagation, they haven’t taken any step to make sure the agent would provide a perfect counteroffer during negotiation. Here, we offered a method to deal with each of these issues.

**Graph Neural Networks (GNN) :** Graphical structure encoding has been successfully accomplished using hierarchical graph pooling-based encoders (Zhang et al., 2019). GNN-based encoders offer better expressive capabilities than HMM (Hid-

den Markov Model) and FST-based encoders and can be trained by optimizing the downstream loss. As they can be interpreted based on observed explicit sequences, they can also improve the model’s interpretability (Tu et al., 2020; Norcliffe-Brown et al., 2018). Graphs have been employed in dialogue systems to direct response selections and dialogue policy. However, rather than composing dialogue strategies on-the-fly, they have been used to encode external knowledge (Tuan et al., 2019; Zhou et al., 2018) or speaker information (Ghosal et al., 2019). The first person to combine GATs with hierarchical pooling was Joshi et al. (2021), who also used a conversation system to acquire practical dialogue techniques. Their model can be plugged into other models as an explicit sequence encoder, unlike earlier works. In order to provide more effective and appropriate negotiation responses, we have drawn inspiration from their concept and employed GAT embeddings to learn the perfect sequencing of negotiation strategies.

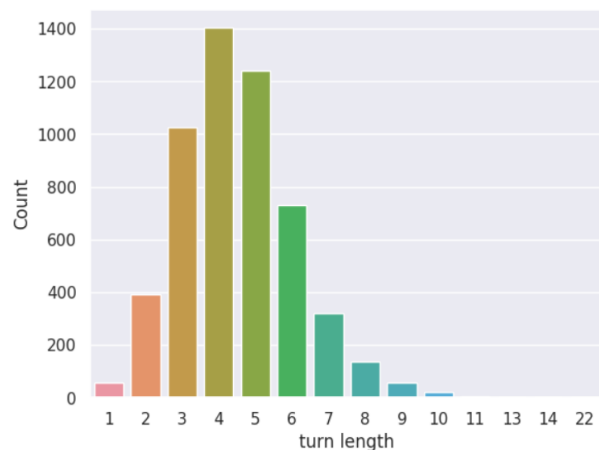


Figure 2: Turn length distribution of CraigListBargain dataset after the removal of faulty utterances

### 3 Dataset

For our training and evaluation purpose, we have chosen CraigListBargain (He et al., 2018) dataset. Previous negotiating datasets were gathered in the context of games. For instance, Asher et al. (2016) gathered conversation records from players of Settlers of Catan online. Two persons were asked to split a collection of hats, books, and balls by Lewis et al. (2017). While these games are useful for grounding and evaluation, they limit the dialogue domain and the linguistic richness.

In CraigListBargain, two agents are given the roles of a buyer and a seller and instructed to nego-

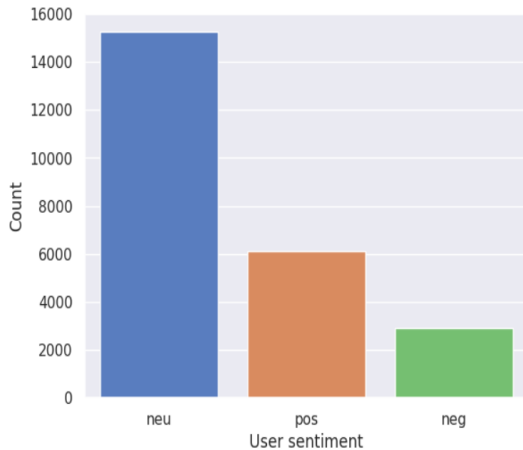


Figure 3: Sentiment distribution of CraigListBargain fetched from VADER (Hutto and Gilbert, 2014)

Annotations	Description,
	User utterances,
	Agent responses,
	User personality,
	Negotiation strategies,
	Dialogue act
# of Negotiation strategies	21
# of Dialogue act	14
# of dialogues	5388
Avg. # of turns	4.97
Avg. # of tokens per turn	12.2

Table 1: Statistics of CraigListBargain dataset

tiate the price of an item up for sale on Craigslist with a description and images in order to promote more open-ended, realistic bargaining. The listing price is displayed to both agents, just as on the actual platform. Additionally, they recommend a target price that is specific to the buyer. The agents take turns openly conversing. Any time, either agent may submit an offer price, which the partner may accept or reject. Additionally, agents have the choice to leave, in which case there is no agreement and the task is completed.

On Amazon Mechanical Turk, they have gathered a good number of conversations between individuals. Some conversational turns are simply blank or include just insignificant tokens. These turns have been dropped since they could throw the model off and result in ridiculous responses. Statistics for the dataset are provided in Table 1 following this filtering. Compared to preceding datasets, CraigListBargain has a sufficient number of turns per dialogue (Shown in Fig 2) and a greater range of utterances. Workers were also urged to adorn the goods and bargain for extras like free delivery or pick-up. This incredibly realistic situation

encourages deeper conversations.

**Sentiment annotation :** For this task, we used the rule-based sentiment analysis tool VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto and Gilbert, 2014), which is created to examine the emotional tone of text documents. This tool accepts a statement as input and outputs the probability distribution across a sentiment space with three possible sentiments (positive, negative, and neutral). We determine the user’s emotion at that turn by taking the sentiment with the greatest probability value from this output. The sentiment distribution of this dataset has been shown in Fig. 3.

## 4 Proposed Methodology

This section gives a detailed description of how we created our graph and integrated it into GPT-2 to enhance contextualization.

### 4.1 Graph Neural Network

This section reviews Graph Attention Networks (GATs) (Velickovic et al., 2017; Li et al., 2021), which has been used in this article. With a node set  $V$  made up of  $N$  nodes  $v_i$  and an edge set  $E$  made up of all connections between nodes, a weighted undirected graph is defined as  $G = (V, E)$ . A binary symmetric adjacency matrix  $N \times N$  is defined as  $S$ , where  $[S]_{ij} = 1$  if  $(v_i, v_j) \in E$  and 0 otherwise. Each node  $v_i$  has feature vectors  $x_i \in R^F$  attached to it. These are compiled into  $N \times F$  dimensional matrices called  $X$ , where  $F$  stands for the input feature size.

Keep in mind that  $SX$  (dot product of matrix  $S$  and  $X$ ) is similar mathematically to transmitting each graph node’s attributes to its neighbours. In this manner,  $S^k X = S(S^{k-1} X)$  is similar to  $k$  rounds of feature swaps with neighbours. As seen in Fig. 5,  $k = 0$  represents self-connection, whereas  $k > 0$  combines characteristics from  $k$  nodes away.

An input  $X \in R^{N \times F}$  is transformed by a GAT layer into an output  $G(X) \in R^{N \times G}$ . As illustrated in Fig. 5, each  $K$ -hop GAT layer is made up of  $P$  attention heads  $A^{(p)}$  that combine  $k = 0, \dots, K - 1$  cycles of feature aggregation over the graph as

$$A^{(p)}(X; S) = \sum_{k=0}^{K-1} (E \odot S)^k X A_k^{(p)} \quad (1)$$

$$G(X) = \frac{1}{P} \sum_{p=1}^P \sigma[A^{(p)}(X; S)],$$

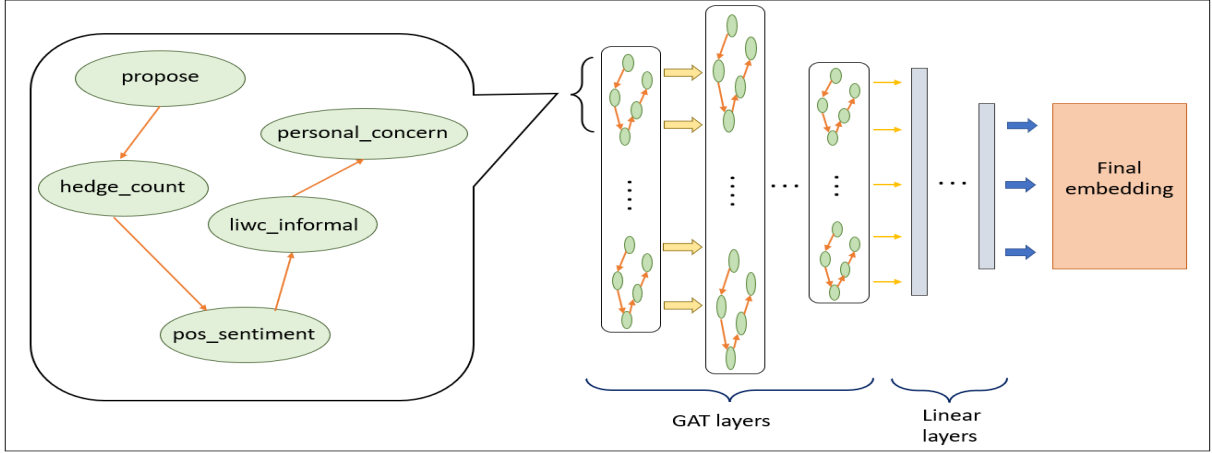


Figure 4: A dummy representation of our Graph Neural Network for negotiation strategies (a similar kind of architecture has been followed for dialogue acts)

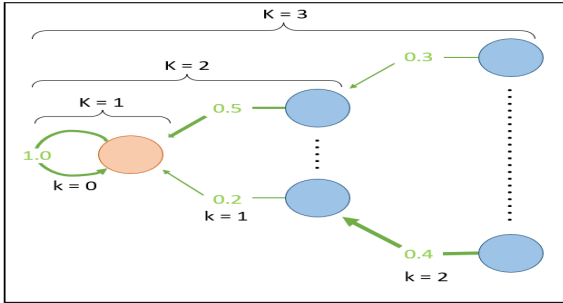


Figure 5: An example of a GAT. Self-connection is represented by  $k = 0$ , whereas passing other nodes' characteristics to the node being evaluated is represented by  $k \geq 1$ . The values on the links are attention values, which weight the passing features

where the  $\{A_k^{(p)}\}_{k=0}^{K-1}$  are  $R^{F \times G}$  linear feature transformations, whereas  $\sigma(\cdot)$  is a non-linear activation function. The values of the  $N \times N$  attention matrix  $E$  are calculated over  $X$  as

$$[E]_{ij} = \frac{\exp(\text{LeakyReLU}(e_{ij}))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(e_{ik}))} \quad (2)$$

$$e_{ij} = (x^i)^T Q^{(p)} x^j,$$

where  $N_i$  are nodes that are adjacent to node  $v_i$ , and  $Q^{(p)}$  are trainable  $F \times F$  matrices that are used to calculate the attention. This is how, assigning dynamic weights to graph edges based on the input node attributes allows a GAT layer to selectively aggregate features.

GATs are built up from a series of  $L$  GAT layers  $G_l$ , each with its own multi-headed graph attention mechanisms  $A_l^p$ . A set of input features  $X_t^{(0)}$  are transformed by the GAT into a set of output features

$X_t^{(L)}$  at time  $t$  as

$$X_t^{(l)} = G_l(X_t^{(l-1)}) \quad \text{for } l = 1, \dots, L \quad (3)$$

Note that, in this study, we conducted experiments using a range of output dimensions, and the findings for these experiments can be found in Table 6 in the appendix.

## 4.2 Graph Formation

We have developed two separate graphs for negotiation strategies and dialogue acts. For negotiation strategies, we have defined the graph as  $G_{ns} = (V_{ns}, E_{ns})$ .  $V_{ns}$  is the set of vertices or nodes where each node represents different negotiation strategies present in the dataset. Finally, we draw a directed edge ( $e_{ns}^{i,j} \in E_{ns}$ ) between node  $v_{ns}^i$  and node  $v_{ns}^j$  if these nodes are appearing in a consecutive manner in any turn of the dialogue. Following the same manner we define the graph for dialogue acts ( $G_{da}$ ). For both instances, the overall Graph neural network is composed of a number of GAT layers followed by a number of linear layers. A miniature representation of our GNN for Negotiation strategies has been shown in Fig. 4. On the left side of this image, we can see a clearer view of one of the graphs where different negotiation strategies ('propose', 'hedge\_count', 'pos\_sentiment', 'liwc\_informal' and 'personal\_concern') represents separate nodes and they are connected according to their occurrence on that turn.

## 4.3 Nego-GAT Architecture

We mostly adhered to the workflow proposed by Yang et al. (2021), incorporated the previously described GNN, and added the sentiment token as

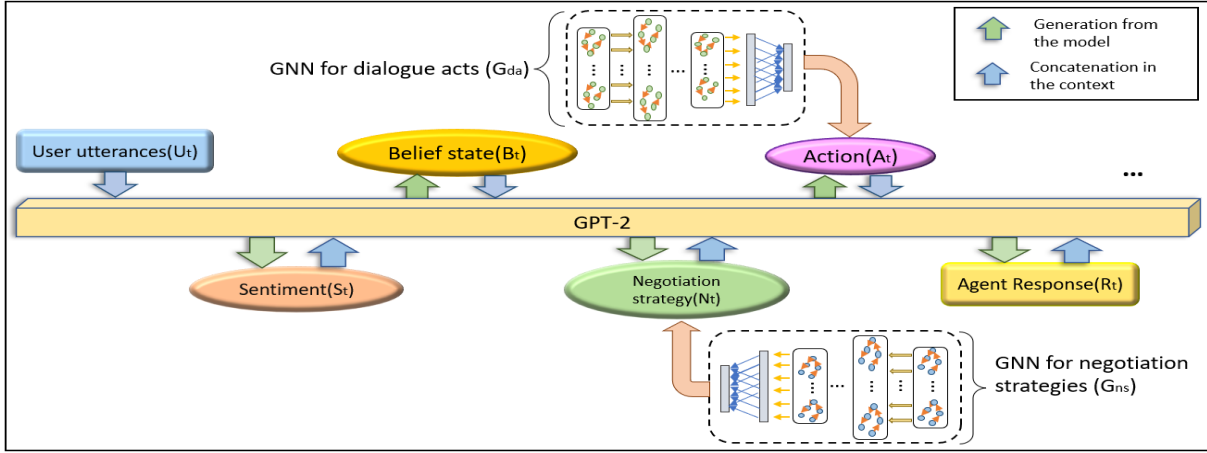


Figure 6: The overall flow of our architecture. We begin our flow from the very left, with the very first utterance given by the user. Following that, our model continues to extract/generate numerous components (sentiment, belief state, negotiation strategies, dialogue action, and responses) and concatenates them into the context for the upcoming turns

```

<eos_u>hi i am interested in your board !<eos_u> <sos_s>neu<eos_s> <sos_b><eos_b>
<sos_ns>[third_person_singular_seller] [-3673] [1610] [-5781] [5380] [7361]
[number_of_diff_dic_pos] [-3666] [1604] [-5774] [5370] [7353]
[number_of_diff_dic_neg] [-3688] [1633] [-5776] [5376] [7355] <eos_ns><sos_a> inquiry
554 673 221 469 110 <eos_a><sos_r> hi do you have any questions ?<eos_r><sos_u> if i
picked it up at your location would you accept <price>_0.2 ?<eos_u> <sos_s>pos<eos_u>
<sos_b>Price_ratio=0.2<eos_b> <sos_ns>[first_person_singular_count_seller] [-3672]
[1607] [-5802] [5395] [7385] <eos_ns> <sos_a>counterprice 523 680 278 463 112
<eos_a>...

```

Figure 7: An example of our overall context. Each component has been encased in different sets of tokens

additional information into the context. GPT-2 is intended to simply obtain the whole context ( $C_t$ ) and predict the next word by computing the conditional probabilities for every word over the entire lexicon ( $V$ ), as illustrated below :

$$W_t = \arg \max_{w \in V} P(w|C_t) \quad (4)$$

Utilizing this structure to our advantage, we trained our model using a self-supervised learning method to be able to generate each and every context component on its own. Our overall work-flow looks like this:

1. At turn  $t = 0$ , the user makes their initial utterance,  $U_0$ .
2. Based on this  $\{U_0\}$  the model predicts the user sentiment  $S_0$  and concatenates it into the context.
3. The model will now derive the belief state,  $B_0$ , based on this  $\{U_0, S_0\}$ . The “price” is the sole feasible belief state in our dataset.

4. The model must select the appropriate collection of negotiating strategies ( $N_0$ ) based on  $\{U_0, S_0, B_0\}$  while maintaining their order. At this point, the previously defined GNN architecture is put into action. At every turn, our model chooses the set of negotiation strategies one by one. After choosing the first negotiation strategy ( $ns_0^{(0)}$ ) of this very first turn we get the respective graph embedding vector from our GNN ( $G_{ns}$ ) and concatenate it into the context. Based on this, the model predicts the next negotiation strategy ( $ns_1^{(0)}$ ). Following this, the model chooses the whole set of negotiation strategies ( $N_0 = \{ns_0^{(0)}, ns_1^{(0)}, \dots, ns_n^{(0)}\}$ ) for this turn.

5. This  $N_0$  is also concatenated into the context. The model now chooses the collection of dialogue acts  $A_0$ , depending on  $\{U_0, S_0, B_0, N_0\}$ . We anticipate  $A_0$  using the same method as for negotiation strategies, using the relevant graph ( $G_{da}$ ) made specifically for dialogue actions. The model guesses the first action  $da_0^{(0)}$ , gets the graph embedding vector, concatenates it to the context, and then moves on to the next dialogue act prediction. This way, finally it predicts the whole set of actions ( $A_0 = \{da_0^{(0)}, da_1^{(0)}, \dots, da_n^{(0)}\}$ ) for this turn.
6. We concatenate all these components in order to form the final context for response generation. So our final context is a stack of the

overall history consisting of user utterance ( $U_0$ ), sentiment ( $S_0$ ), belief state ( $B_0$ ), negotiation strategy ( $N_0$ ), and dialogue act ( $A_0$ ). Based on this, our model provides response  $R_0$  and this completes the very first turn.

Up until the conversation’s conclusion, this flow has been maintained. Each turn, the elements from every prior turn are concatenated to construct the final context. One example of our context and our overall architecture are shown in Fig. 7 and Fig. 6 respectively.

## 5 Experimental Results

In this section, we present all the findings from our research.

### 5.1 Baselines and Implementation Details

**Nego-GAT** is our proposed work. We compare Nego-GAT’s performance to that of **Dialog-Graph** (Joshi et al., 2021), the most comprehensive model for the negotiating task now available, which also employs GAT to store dialogue actions and strategy sequences. The FST-enhanced hierarchical encoder-decoder model (FeHED) (Zhou et al., 2019b), which was the previous state-of-the-art model, was also used to compare our agent’s performance. Furthermore, we conduct a comparative study utilizing three different encoding techniques for negotiation strategies: **HED**, **HED+RNN**, and **HED+Transformer**. HED completely disregards both, in contrast to HED+RNN and HED+Transformer, which employ RNN and Transformers (Vaswani et al., 2017) to encode the strategy and dialogue act information, respectively.

Our model has been implemented at the session level using DistilGPT2 (Sanh et al., 2019), a distilled version of GPT-2, and HuggingFace’s Transformers (Wolf et al., 2019). This shows that the model has learned to create or predict the next word depending on the current word and has taken feedback from the entire dialogue into account. In CraigListBargain dataset we have multiple negotiation strategies and dialogue acts on each single turn and they are well mixed up. Keeping this in mind, dividing the dataset randomly into 3:1 ratio (train:test) we trained the model across 15 epochs on the training data using a temperature of 0.7, cross-entropy function as our loss function, and AdamW as our optimizer. Additionally, we defined the size of our embedding space, which is 5 (for both negotiation strategy and dialogue act), and

chose the ideal GAT and linear layer configuration for our GNN, which is 3 GAT and 2 linear layers in this scenario, using an ablation study (See Appendix A for detailed results).

### 5.2 Discussion of Results

We examine our model’s propensity for predicting both dialogue actions and negotiating strategies (internal evaluation) as well as its ability to elicit replies (external assessment). We examine dialogue generation in three different contexts: first, without negotiation strategy embeddings; second, without dialogue act embeddings; and third, with both embeddings present. Each outcome is discussed in extensive detail further down.

**Negotiation strategy & dialogue act prediction :** Apart from generating dialogues our model is capable of choosing the necessary negotiation strategies and dialogue actions at each turn, which are nothing but a classification tasks. There aren’t many full models for generating negotiation discourse that employ distinct modules to carry out these responsibilities. To assess our model’s suitability for the task, we compare its performance with that of similar cutting-edge models. Due to the unbalanced class distributions of these two components (negotiation strategy and dialogue act), we only take the F1 score into account when assessing the models. Table 2 reports the results of the respective models. The table demonstrates that GPT-2 performs nearly on par with the earlier models even without employing graph embeddings (see 4th row of the table), but the addition of graph embeddings (last row of the table) increases their performance significantly.

**Dialogue generation :** Our goal was to produce effective negotiation talks based on a context made up of the sentiment, belief state, negotiation strategy, and dialogue act. To determine how closely the produced replies match the gold human responses, we have utilized a number of automated assessment metrics, including the BLEU (BiLingual Evaluation Understudy) (Papineni et al., 2002), precision, recall, and F1 score. The results are provided in Table 4. As we have already shown, the usage of graph embeddings greatly aided in the selection of the dialogue act and negotiating approach. Now, we can see how it influences dialogue production as well. From Table 4 it is very clear that choosing the right set of negotiation strategies eventually helps our model to generate dialogues closer to our

Model	Negotiation strategy			Dialogue act		
	F1 score			F1 score		
	Macro	Micro	Weighted	Macro	Micro	Weighted
FeHED [(Zhou et al., 2019b)]	12.3	15.7	17.8	14.2	18.2	25.4
HED+RNN	15.1	20.5	22.7	17.7	21.3	29.4
HED+Transformer	17.8	22.1	24.1	18.8	22.6	31.3
DialoGraph [(Joshi et al., 2021)]	18.1	23.7	26.9	20.1	24.6	32.7
GPT-2 without Graph embedding	18.3	23.5	27.4	21.7	25.5	33.3
Nego-GAT(our model)	<b>20.3</b>	<b>26.5</b>	<b>33.2</b>	<b>24.3</b>	<b>29.1</b>	<b>37.5</b>

Table 2: Performance comparison of different baselines and our model for the task of negotiation strategy and dialogue act prediction

Model	Persuasive	Coherent	Natural	Understandable	Sale Price Ratio	Avg words/turn
HED	2.50	2.50	3.50	2.50	-2.13	4.25
FeHED (Zhou et al., 2019b)	3.30	3.75	3.70	3.69	0.25	5.76
HED+RNN	2.81	3.27	3.36	3.27	-3.68	3.61
HED+Transformer	3.50	3.50	3.70	3.40	-0.07	4.36
DIALOGRAPH (Joshi et al., 2021)	3.58	3.94	3.75	3.70	0.49	5.84
Nego-GAT (without any embedding)	3.86	3.95	<b>4.12</b>	<b>4.50</b>	0.52	6.22
Nego-GAT (Our final model)	<b>3.98</b>	<b>4.00</b>	<b>4.12</b>	<b>4.50</b>	<b>0.56</b>	<b>8.54</b>

Table 3: Human evaluation results to assess the qualitative aspects of the generated responses

Model	BLEU	Precision	Recall	F1 score
HED	10.9	10.0	9.3	7.8
FeHED (Zhou et al., 2019b)	12.1	12.6	12.0	11.0
HED+RNN	13.1	12.5	12.9	11.5
HED+Transformer	15.2	14.7	14.5	13.7
DialoGraph (Joshi et al., 2021)	15.2	14.9	14.8	14.1
Nego-GAT without any embedding	16.9	16.5	16.1	<b>15.8</b>
Nego-GAT with strategy embedding	17.7	<b>17.5</b>	16.9	15.2
Nego-GAT (Our final model)	<b>17.9</b>	<b>17.5</b>	<b>17.1</b>	15.6

Table 4: The results of automatic evaluation to measure how closely the generated responses correspond to the gold standard

standard gold responses.

**Human evaluation:** To gauge how well these models perform qualitatively, we have also used human evaluations (See Table 3). As part of the human evaluation, we calculated the average number of words per turn, the average sales price ratio on which the deal was finalized, and four user ratings to assess four qualitative parameters (persuasive, coherent, natural, and understandable) in the generated responses on a scale of 1 to 5. In essence, these evaluations assess the following characteristics of the produced responses:

1. **Persuasive** : how well it persuades the user to accept a greater offer.
2. **Coherent** : How logically related or coherent the answers are.
3. **Natural** : To ascertain if the generated responses are more human-like or not.
4. **Understandable** : It is to evaluate the grammar and clarity of the responses.

The outcomes of both tables (Table 3 and Table 4) show that the replies we developed are more precise, fluid, and calming. Even the increase in the average sales price ratio and average words per turn demonstrates that in addition to generating rational and accurate replies, it performs extremely well in negotiations. An example, of comparing the outcomes of different modules is shown in Appendix B. Additionally, a full generation pipeline has been shown in Appendix C.

### 5.3 Error Analysis

Although our model is capable of carrying out multiple tasks by itself, it is unable to manage the counter price it is proposing. In Fig. 9 (in Appendix) one example of such scenario is shown. Here, we can see that in the very first turn, our model makes an offer of price 6, but at the very next turn, instead of lowering its offer, it raises it to price 10. As a result, the deal did not take place. Our model has only been trained to predict the next token based on the present token. As a result, it is unable to compare the digits to their values. We will seek for a solution to this issue during our future study.

## 6 Conclusion and Future Works

This study demonstrates how combining GNN with the comprehensive pre-trained language model GPT-2 aids in selecting the appropriate set of strategies and dialogue actions, thereby improving the



efficacy of negotiating discussions. Because of graph embeddings, which provide information on the sequence in which negotiation strategies were used at the ground truth, our model can successfully imitate these strategies in a real-world context, keeping the generated dialogues fluid and appealing.

## 7 Limitations

Despite all the above-mentioned benefits, our approach has certain drawbacks as well. Only the English language can be used to communicate with this model, and it is very reliant on the type of data provided to it during training. For instance, it cannot effectively bargain with a consumer who wants to purchase some flowers, since conversations related to selling flowers were not included in the dataset. In addition, even if this model is producing counter prices that are higher than those of the older models, it may not always be effective at closing deals. To do that, we must create counter-proposals based on the person’s personality feature or some other calculated tactic. In the future, we’ll look for answers to these issues and create a model for this task that is more potent.

## 8 Ethics Statement

We used widely used, publicly accessible datasets to model negotiations. Without violating any copyright concerns, we complied with the dataset’s regulations. Additionally, we employed another openly accessible and publicly available tool to obtain the sentiment annotations on this dataset. It is necessary to consider an ethical aim while creating bargaining conversational AI. Only the product price is subject to negotiation in our utilized dataset. We made an effort to create a flawless model that could negotiate deals with user parties with ease and composure and, in our opinion, can make a significant influence on the field of online exchange and marketing. Last but not least, because generative models lack context-sensitive information, they may produce uninformative statements. As a result, it is necessary to model knowledge grounding or fact-verification. We’ll use the findings of this investigation in our future works.

## Acknowledgements

This paper presents research findings from the project “Conversational Agents with Negotiation and Influencing Ability” (Project No.

IITP/2023/814), sponsored (faculty research grant) by Accenture Pvt Ltd. We are grateful to Accenture Pvt Ltd and the research team for collaborating on the research problem with huge real-world implications.

## References

- Nicholas Asher, Julie Hunter, Mathieu Morey, Farah Benamara, and Stergos Afantenos. 2016. Discourse structure and dialogue acts in multiparty dialogue: the stac corpus. In *10th International Conference on Language Resources and Evaluation (LREC 2016)*, pages 2721–2727.
- Paweł Budzianowski and Ivan Vulić. 2019. Hello, it’s gpt-2—how can i help you? towards the use of pre-trained language models for task-oriented dialogue systems. *arXiv preprint arXiv:1907.05774*.
- Ssu Chiu, Maolin Li, Yen-Ting Lin, and Yun-Nung Chen. 2022. Salesbot: Transitioning from chit-chat to task-oriented dialogues. *arXiv preprint arXiv:2204.10591*.
- Marian Friestad and Peter Wright. 1994. The persuasion knowledge model: How people cope with persuasion attempts. *Journal of consumer research*, 21(1):1–31.
- Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. Dialoguecn: A graph convolutional neural network for emotion recognition in conversation. *arXiv preprint arXiv:1908.11540*.
- He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. *arXiv preprint arXiv:1808.09637*.
- Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.
- Rishabh Joshi, Vidhisha Balachandran, Shikhar Vashishth, Alan Black, and Yulia Tsvetkov. 2021. Dialograph: Incorporating interpretable strategy-graph networks into negotiation dialogues. *arXiv preprint arXiv:2106.00920*.
- George Larionov, Zachary Kaden, Hima Varsha Dureddy, Gabriel Bayomi T Kalejaiye, Mihir Kale, Srividya Pranavi Potharaju, Ankit Parag Shah, and Alexander I Rudnicky. 2018. Tartan: A retrieval-based socialbot powered by a dynamic finite-state machine architecture. *arXiv preprint arXiv:1812.01260*.
- Mike Lewis, Denis Yarats, Yann N Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or no deal? end-to-end learning for negotiation dialogues. *arXiv preprint arXiv:1706.05125*.

- Jiwei Li, Michel Galley, Chris Brockett, Georgios P Spithourakis, Jianfeng Gao, and Bill Dolan. 2016. A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155*.
- Jiwei Li, Will Monroe, Tianlin Shi, Sébastien Jean, Alan Ritter, and Dan Jurafsky. 2017. Adversarial learning for neural dialogue generation. *arXiv preprint arXiv:1701.06547*.
- Qingbiao Li, Weizhe Lin, Zhe Liu, and Amanda Prorok. 2021. Message-aware graph attention networks for large-scale multi-robot path planning. *IEEE Robotics and Automation Letters*, 6(3):5533–5540.
- Yu Li, Kun Qian, Weiyan Shi, and Zhou Yu. 2020. End-to-end trainable non-collaborative dialog system. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8293–8302.
- Bing Liu and Ian Lane. 2018. End-to-end learning of task-oriented dialogs. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 67–73.
- Will Norcliffe-Brown, Stathis Vafeias, and Sarah Parisot. 2018. Learning conditioned graph structures for interpretable visual question answering. *Advances in neural information processing systems*, 31.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.
- Richard E Petty and John T Cacioppo. 1986. *The elaboration likelihood model of persuasion*. Springer.
- Shuang Qiu and Kang Zhang. 2021. Learning personalized end-to-end task-oriented dialogue for fast and reliable adaptation. In *2021 International Conference on Digital Society and Intelligent Systems (DSInS)*, pages 62–66. IEEE.
- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.
- Aritra Raut, Subrata Das, Abhisek Tiwari, Sriparna Saha, Anutosh Maitra, Roshni Ramnani, and Shubhashis Sengupta. 2022. Introducing multi-modality in persuasive task oriented virtual sales agent. In *International Conference on Neural Information Processing*, pages 543–555. Springer.
- Aritra Raut, Abhisek Tiwari, Subrata Das, Sriparna Saha, Anutosh Maitra, Roshni Ramnani, and Shubhashis Sengupta. 2023. Reinforcing personalized persuasion in task-oriented virtual sales assistant. *Plos one*, 18(1):e0275750.
- Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
- Alessandro Sordani, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Margaret Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015. A neural network approach to context-sensitive generation of conversational responses. *arXiv preprint arXiv:1506.06714*.
- Kai Sun, Seungwhan Moon, Paul Crook, Stephen Roller, Becka Silvert, Bing Liu, Zhiguang Wang, Honglei Liu, Eunjoon Cho, and Claire Cardie. 2020. Adding chit-chat to enhance task-oriented dialogues. *arXiv preprint arXiv:2010.12757*.
- Abhisek Tiwari, Tulika Saha, Sriparna Saha, Shubhashis Sengupta, Anutosh Maitra, Roshni Ramnani, and Pushpak Bhattacharyya. 2021. A dynamic goal adapted task oriented dialogue agent. *Plos one*, 16(4):e0249030.
- Abhisek Tiwari, Tulika Saha, Sriparna Saha, Shubhashis Sengupta, Anutosh Maitra, Roshni Ramnani, and Pushpak Bhattacharyya. 2022. A persona aware persuasive dialogue policy for dynamic and co-operative goal setting. *Expert Systems with Applications*, 195:116303.
- Ming Tu, Kevin Huang, Guangtao Wang, Jing Huang, Xiaodong He, and Bowen Zhou. 2020. Select, answer and explain: Interpretable multi-hop reading comprehension over multiple documents. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 9073–9080.
- Yi-Lin Tuan, Yun-Nung Chen, and Hung-yi Lee. 2019. Dykgchat: Benchmarking dialogue generation grounding on dynamic knowledge graphs. *arXiv preprint arXiv:1910.00610*.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. 2017. Graph attention networks. *stat*, 1050(20):10–48550.
- Xuwei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. *arXiv preprint arXiv:1906.06725*.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2019. Huggingface’s transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.

- Yunyi Yang, Yunhao Li, and Xiaojun Quan. 2021. Ubar: Towards fully end-to-end task-oriented dialog system with gpt-2. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14230–14238.
- Steve Young. 2006. Using pomdps for dialog management. In *2006 IEEE Spoken Language Technology Workshop*, pages 8–13. IEEE.
- Zhen Zhang, Jiajun Bu, Martin Ester, Jianfeng Zhang, Chengwei Yao, Zhi Yu, and Can Wang. 2019. Hierarchical graph pooling with structure learning. *arXiv preprint arXiv:1911.05954*.
- Hao Zhou, Tom Young, Minlie Huang, Haizhou Zhao, Jingfang Xu, and Xiaoyan Zhu. 2018. Commonsense knowledge aware conversation generation with graph attention. In *IJCAI*, pages 4623–4629.
- Yiheng Zhou, He He, Alan W Black, and Yulia Tsvetkov. 2019a. A dynamic strategy coach for effective negotiation. *arXiv preprint arXiv:1909.13426*.
- Yiheng Zhou, Yulia Tsvetkov, Alan W Black, and Zhou Yu. 2019b. Augmenting non-collaborative dialog systems with explicit semantic and strategic dialog history. *arXiv preprint arXiv:1909.13425*.

## A Ablation Study

In order to find the best GAT and linear layer combination for our GNN, we currently fixed the embedding size at 3 and experimented with several combinations. After getting the embeddings (for both dialogue acts and negotiation strategies) from each of these GNNs we have trained our model for 5 epochs each and noted their performances in Table 5. Following these experiments we chose 3 GAT - 2 Linear as our final combination of our GNN. After that, we conducted another investigation to determine the ideal embedding size for our task. Using the previous determined GNN structure we trained the model for 20 epochs and noted the result in Table 6. From this table it is quite clear that it does not provide a lot of information concerning upcoming strategies for size 3. On the other hand, the context typically becomes too vast when the embedding size is increased to 7. Because of this, the majority of contexts are larger than the embedding size of GPT-2, losing other significant context elements in the process. Using an embedding size of 5, it appears we can achieve a flawless trade-off between these two scenarios.

## B Conversations Comparison

Outcomes of DialoGraph and our model (with and without graph embeddings) is shown in Fig. 8. The

results clearly demonstrate that DialoGraph (Joshi et al., 2021) is effective but not very attractive during negotiations. While our models are not just producing greater counter offers, the average number of words at each turn is also significantly higher. Another thing to note is that although our model first offered a lesser counter offer, it later chose the appropriate strategy and concluded the sale at a higher price which can actually be effective in real life scenarios.

## C Example of a Complete Conversation

Nego-GAT is capable of producing sentiment, belief states, negotiation strategies, dialogue actions and finally the responses. An example of this whole generation flow has been shown in Fig. 10. The graph embedding of each of the negotiation strategies (green in the picture) and dialogue acts (purple in the image) is also displayed for clarification at each step.

GAT layer	Linear layer	BLEU	Precision	Recall	F1 score
1	2	3.76	1.40	1.79	1.57
	3	3.41	1.36	1.65	1.49
2	2	5.01	1.97	2.23	1.09
	3	4.79	1.76	2.05	1.89
3	2	<b>5.89</b>	<b>2.45</b>	<b>3.01</b>	<b>2.70</b>
	3	5.76	2.23	2.87	2.51

Table 5: Result of ablation study to determine the ideal GAT-linear layer combination for our GNN

Model	Embedding size	BLEU	Precision	Recall	F1 score
Nego-GAT with strategy embedding	3	19.3	14.76	14.89	13.10
	5	21.6	18.41	18.39	16.74
	7	20.1	16.22	16.78	14.14
Nego-GAT with both embeddings	3	19.1	14.92	14.78	13.34
	5	<b>21.9</b>	<b>18.34</b>	<b>18.45</b>	<b>18.39</b>
	7	20.1	16.43	16.67	13.93

Table 6: Result of ablation study to determine the optimal size of graph embeddings

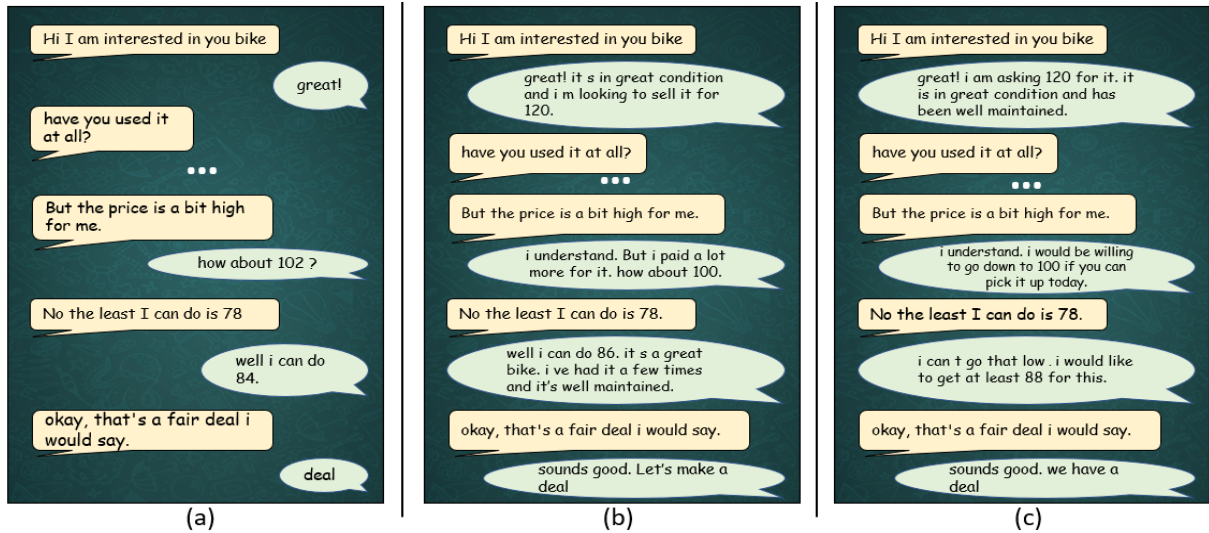


Figure 8: Model outputs. (a) the output of DialoGraph (Joshi et al., 2021), (b) output of Nego-GAT w/o any embeddings, (c) output of Nego-GAT (our final model) (In every conversation, the left side dialogues (in rectangles) are created by the user, while the right side dialogues (in ellipsoidal curves) are generated by the agent.)



Figure 9: An example of an unsuccessful conversation generated by our model. The left side dialogues (in rectangles) are created by the user, while the right side dialogues (in ellipsoidal curves) are generated by the agent

---

### Marked price of the Item : 10 USD

---

User : hello i am interested in this calculator. is this old?

Generated sentiment : pos

Generated belief state :

Negotiation strategy : [pos\_sentiment] [-3671] [1608] [-5793] [5387] [7374] [third\_person\_singular] [-3673] [1610] [-5781] [5380] [7361] [number\_of\_diff\_dic\_pos] [-3666] [1604] [-5774] [5370] [7353]

Generated action : unknown 533 896 289 321 100

Generated response : it is old but works great.

---

User : what colour is it ?

Generated sentiment : neu

Generated belief state :

Negotiation strategy : [pos\_sentiment] [-3671] [1608] [-5793] [5387] [7374] [third\_person\_singular] [-3673] [1610] [-5781] [5380] [7361] [number\_of\_diff\_dic\_pos] [-3666] [1604] [-5774] [5370] [7353] [number\_of\_diff\_dic\_neg] [-3688] [1633] [-5776] [5376] [7355] [personal\_concern] [-3668] [1606] [-5797] [5388] [7377]

Generated action : inform 530 860 250 387 102

Generated response : it is dark brown and has a few small scratches on the top. Otherwise it is in great condition.

---

User : okay great. i wanna get it but can you come down a bit on the price? is 4 acceptable?

Generated sentiment : neu

Generated belief state : Price\_ratio-0.4

Negotiation strategy : [pos\_sentiment] [-3671] [1608] [-5793] [5387] [7374] [first\_person\_singular\_count] [-3672] [1607] [-5802] [5395] [7385] [number\_of\_diff\_dic\_pos] [-3666] [1604] [-5774] [5370] [7353] [number\_of\_diff\_dic\_neg] [-3688] [1633] [-5776] [5376] [7355] [hedge\_count] [-3685] [1624] [-5776] [5377] [7357] [personal\_concern] [-3668] [1606] [-5797] [5388] [7377]

Generated action : counter\_price 540 872 289 319 102

Generated response : i would be willing to go down to 7.

---

User : okay works for me then! can you please meet me somewhere so i can pick it up from you or can you deliver it to me?

Generated sentiment : neu

Generated belief state :

Negotiation strategy : [seller\_pos\_sentiment] [-3671] [1608] [-5793] [5387] [7374] [first\_person\_singular\_count\_seller] [-3672] [1607] [-5802] [5395] [7385] [third\_person\_singular\_seller] [-3673] [1610] [-5781] [5380] [7361] [number\_of\_diff\_dic\_pos] [-3666] [1604] [-5774] [5370] [7353] [number\_of\_diff\_dic\_neg] [-3688] [1633] [-5776] [5376] [7355] [hedge\_count\_seller] [-3685] [1624] [-5776] [5377] [7357] [personal\_concern\_seller] [-3668] [1606] [-5797] [5388] [7377]

Generated action : unknown 533 896 289 321 100

Generated response : i would be willing to meet you at my location. i will be able to deliver it to you.

---

User : great then we got a deal !

Generated sentiment : pos

Generated belief state :

Negotiation strategy :

Generated action : unknown 533 896 289 321 100

Generated response : great!

---

Figure 10: An example of an overall conversation flow of Nego-GAT (our final model)