A MIXTURE-OF-EXPERT APPROACH TO RL-BASED DIALOGUE MANAGEMENT

Yinlam Chow **Aza Tulepbergenov** **Ofir Nachum** **Dhawal Gupta**

Moonkyung Ryu **Mohammad Ghavamzadeh** **Craig Boutilier**

Google Research
{yinlamchow,atulep,ofirnachum,dhawgupta,mkryu,ghavamza,cboutilier}@google.com

ABSTRACT

Despite recent advancements in language models (LMs), their application to dialogue management (DM) problems and ability to carry on rich conversations remain a challenge. We use reinforcement learning (RL) to develop a dialogue agent that avoids being short-sighted (outputting generic utterances) and maximizes overall user satisfaction. Most existing RL approaches to DM train the agent at the word-level, and thus, have to deal with a combinatorially complex action space even for a medium-size vocabulary. As a result, they struggle to produce a successful and engaging dialogue even if they are warm-started with a pre-trained LM. To address this issue, we develop a RL-based DM using a novel *mixture of expert* language model (MoE-LM) that consists of (i) a LM capable of learning diverse semantics for conversation histories, (ii) a number of *specialized* LMs (or *experts*) capable of generating utterances corresponding to a particular attribute or personality, and (iii) a RL-based DM that performs dialogue planning with the utterances generated by the experts. Our MoE approach provides greater flexibility to generate sensible utterances with different intents and allows RL to focus on conversational-level DM. We compare it with SOTA baselines on open-domain dialogues and demonstrate its effectiveness both in terms of the diversity and sensibility of the generated utterances and the overall DM performance.

1 INTRODUCTION

With the tremendous advancements in natural language understanding and generation, increasing attention has been directed to construct intelligent dialogue agents that can carry out engaging conversations with users. Such interactions can be open-ended, contain different topics, and often involve an underlying task, such as negotiation, information exchange, and recommendation. Therefore, to satisfy the user, a good dialogue agent should not only generate natural responses, but also be capable of pursuing the task's objectives and adapting to the user's feedback on-the-fly.

A standard solution is to train the dialogue agent using behavioral cloning, where the agent is a language model (LM) that imitates the utterances in the training set (Gašić et al., 2011; Fatemi et al., 2016). By leveraging deep neural networks, e.g., RNNs (Sutskever et al., 2014) and Transformers (Vaswani et al., 2017), a LM encodes the conversation to a low-dimensional dialogue state and predicts an utterance, but steering such generation for particular purposes remains an open question. Several works studied ways to fine-tune a LM to generate texts with specific contexts (Ziegler et al., 2019; Ficler and Goldberg, 2017). Other results learned a single steerable LM that is capable of generating utterances for multiple specific intents (Gu et al., 2017; Chen et al., 2018; Subramani et al., 2019; Dathathri et al., 2019). While these LMs produce fluent and relevant responses, it is unclear how to control them to systematically pursue goals during multi-turn dialogue conversations.

Another popular approach is to view dialogue management (DM) as a control problem and use reinforcement learning (RL) to optimize the agent's policy (which is often a LM itself). Using RL for dialogue systems has a long history. Earlier work relies on specific, hand-crafted semantic states (Levin and Pieraccini, 1997; Singh et al., 2002; Walker, 2000) or partially observable belief states (Williams and Young, 2007; Young et al., 2010), in which the agent chooses the best hand-crafted dialogue act at each turn, with the goal of either satisfying the user (Shah et al., 2018), completing the task (Shi and Yu, 2018), or responding to the user's query (Serban et al., 2017a). However, the application of these approaches is limited to problems whose action space can be captured by hand-crafted representations, and they cannot handle complex conversations. On the other hand, more recent approaches use deep learning to extract semantic representations from conversation histories, treat these representations as dialogue belief states, and apply RL to learn a word-level generative DM agent (Jaques et al., 2019; Li et al., 2016; 2017; Shin et al., 2020). However, since there are innumerable possibilities of language utterances, and thus, the action space of the RL problem is extremely large, the agent often performs planning poorly and generates incomprehensible utterances (Zhao et al., 2019). Another issue is that RL only optimizes a scalar reward, while the aforementioned methods often need to optimize for both the quality of the generated utterance, e.g., ease of answering (Li et al., 2016), fluency (Li et al., 2017; 2019), and diversity (Yarats and Lewis, 2018), and the goal, e.g., conversation length (Zhou et al., 2020), user's sentiment (Hancock et al., 2019), and task completion (Verma et al., 2022; Jang et al., 2021). Moreover, defining the reward as weighted combination of these metrics is not ideal, since the hand-picked weights do not often reflect the underlying success criteria.

To address the above issues related to using RL in dialogue management (DM) systems, we propose an RL-based DM agent using a novel *mixture of expert* (MoE) approach. Our MoE approach is based on a mixture of expert language model (MoE-LM), which consists of three main components: **1)** a LM (a probabilistic encoder and a decoder) capable of learning diverse semantics for conversation histories, and as a result generating diverse utterances, which we refer to as the *primitive* LM or $\texttt{LM}_0$, **2)** a number of *specialized* LMs (or *experts*), $\{\texttt{LM}_i\}_{i=1}^m$, that each is constructed using the latent space learned by $\texttt{LM}_0$, but has been trained such that it is capable of generating utterances corresponding to a certain intent or personality, and **3)** an RL-based dialogue manager (DM) that at each turn, given the latent state shared by the experts $\{\texttt{LM}_i\}_{i=0}^m$ and the utterance action(s) they suggest, chooses one among them for the agent to execute. Our MoE-LM can be seen as a special case of hierarchical LMs (e.g., Serban et al. 2017a; Zhao et al. 2019; Saleh et al. 2020), but it is different than them because it learns both the LMs (experts) and the DM. Moreover, the DM in MoE-LM is a policy conditioned on both the latent state and the actions suggested by the experts, and not just the state as it is common in hierarchical RL. The primitive LM ($\texttt{LM}_0$) plays an important role in this model because it learns diverse semantics for conversation histories and allows the agent to generate a wide variety of utterances. This diversity is also shared with the specialized LMs (experts) and gives them flexibility in generating their (more) specialized utterances. Another important feature of MoE-LM is its modularity that facilitates adding and removing specialized LMs (experts). Moreover, this hierarchical architecture allows us to solve an RL problem with much smaller state and action spaces, which is quite important in the quality of the learned policy. Finally, since the candidate utterances are generated by experts with different intents, instead of combining all agent-user signals into a single RL reward, our DM agent can focus on optimizing the specific goal of the conversation task.

We start the paper with a brief introduction of LMs and the use of Markov decision processes (MDPs) in modeling dialogue management problems in Section 2. We then describe the overall architecture of our MoE-LM in Section 3, followed by the detailed implementation of each of its three main components (described in the above paragraph) in Sections 4 to 6. Finally, in Section 7, we demonstrate the effectiveness of our MoE-LM in open-domain dialogues, in terms of both its ability to generate diverse and sensible utterances and its overall DM performance.

2 Preliminaries

Language Models (LMs) In this work, we employ seq2seq LMs to generate the next utterances in a dialogue. We assume access to a dataset of the form $\mathcal{D} = \{(\mathbf{X}^{(k)}, Y^{(k)})\}_{k=1}^{|\mathcal{D}|}$, where each $\mathbf{X} = \mathbf{X}^{(k)}$ is a L-turn conversation history $\mathbf{X} = \{X_l\}_{l=0}^{L-1}$ and Y is its next utterance. We denote by $N_{\mathbf{X}}$, an upper-bound on the length (number of tokens) of each utterance X_l in $\mathbf{X}$.[1] The role of a LM is to predict the probability of the next utterance Y, consisting of N tokens, conditioned on the conversation history $\mathbf{X}$, i.e., $p\big(Y = \{y_n\}_{n=1}^N \mid \mathbf{X}\big)$. In the transformer architecture (**?**), the LM first encodes the conversation history $\mathbf{X}$ using an encoder Φ to a $(L \times N_{\mathbf{X}})$-length sequence of embeddings $\{(z_{l,0}, \ldots, z_{l,N_{\mathbf{X}}-1})\}_{l=0}^{L-1}$, where each $z_{l,n}$ is a vector in the latent space. For notational convenience, we concatenate these embeddings into a single embedding $z \in \mathcal{Z} \subseteq \mathbb{R}^d$ and denote the overall dimension of the latent space as d. In the RNN architecture (Serban et al., 2016), the LM's encoder Φ directly maps the conversation history $\mathbf{X}$ to a latent state $z \in \mathcal{Z} \subseteq \mathbb{R}^d$. In both architectures, the next utterance $\widehat{Y} = \{\widehat{y}_n\}_{n=1}^N$ is sampled token-by-token from the decoder Ψ,

[1] If the actual utterance X_l has fewer tokens than $N_{\mathbf{X}}$, it will be padded by a specific token and masked.

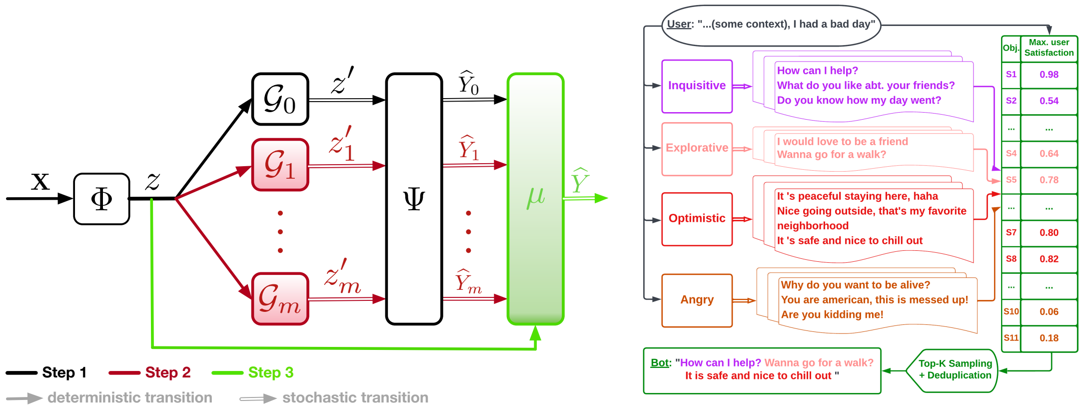


Figure 1: (Left) MoE-LM Architecture. (Right) Sample utterance workflow generated by an MoE-LM trained with Reddit data. Step 1: Φ encodes conversation history. Step 2: $\Psi \circ \mathcal{G}_i, \forall i$, generate candidate bot utterances. Step 3: μ selects the bot response by Q-score ranking & post-processing.

i.e., $\widehat{Y} \sim \Psi\big(\cdot \mid z\big) = \prod_{n=1}^{N} \Psi\big(\widehat{y}_n \mid \widehat{y}_0, \ldots, \widehat{y}_{n-1}; z\big)$, where $\widehat{y}_0$ is a fixed initial (start-of-sentence) token (Chien and Kuo, 2019), and the latent state is denoted as $z = \Phi(\mathbf{X})$.[2]

Markov Decision Processes (MDPs) have been used to model dialogue management problems in a variety of settings (Li et al., 2016; Asadi and Williams, 2016; Jaques et al., 2019). In such MDPs, denoted by $\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, r, s_0, \gamma)$, the state space $\mathcal{S}$ represents the tokenized conversation history and the initial state $s_0 \in \mathcal{S}$ is the initial user's query. The action space $\mathcal{A}$ is also the tokenized language space with each action $a \in \mathcal{A}$ being the agent's next utterance (which is a fixed-length, $N_{\mathbf{X}}$, sequence of tokens). The transition kernel P models the user's response to the action taken by the agent (bot). Finally, the reward function r measures the user's satisfaction. In these MDPs, we can think of the entire LM as a policy that maps conversation histories to next utterances, and solve them by finding a policy π^* with maximum expected discounted return, i.e., $\pi^* \in \arg\max_\pi J_\pi := \mathbb{E}[\sum_{k=0}^{\infty} \gamma^t r_t \mid P, s_0, \pi]$. Note that the size of the tokenized state and action spaces grow exponentially with the size of the vocabulary. This makes it intractable to solve the MDP even for a medium-size vocabulary. As a result, it would quite desirable to develop a novel MDP paradigm that is more amendable to RL-based DM systems.

3 Mixture of Experts (MoE) Language Model

We start by explaining how a MoE language model (MoE-LM) can enrich the bot's utterances and improve the overall performance of the DM. While our approach is applicable to any DM system, we use open-domain dialogue (Sankar et al., 2019) as a running example to show how MoE-LM-based agents can improve user satisfaction measured by an improvement on a sentiment or engagement. Intuitively a good DM agent should possess different behaviors (e.g., inquisitive, explorative, relevant, soothing, empathetic, complimentary, provoking) and swiftly decide which intent to use to pivot a conversation, build rapport, pique the user's interests, improve their mood, etc. To achieve this goal, we require the LM to have a language representation (primitive discovery) that captures different semantics, in order to encode different conversations and avoid generating dull and repetitive responses. We also need a machinery (expert construction) to embed different intents into sub-models of this LM, so that they can behave accordingly when prompted, and respond efficiently. Finally, with various candidate utterances available, the DM module of this LM should understand the current level of user satisfaction and determine which response is the most appropriate. Motivated by these observations, we construct our MoE-LM in three steps as shown in Figure 1. We give the main idea behind each step here and leave their detailed descriptions to Sections 4, 5, and 6.

Step 1: Primitive Discovery. We first employ the dataset $\mathcal{D}$, introduced in Section 2, and learn a language model $\mathtt{LM}_0 = (\Phi, \mathcal{G}_0, \Psi)$ consisting of a *stochastic encoder* (i.e., an encoder Φ and a latent space distribution $\mathcal{G}_0$ that maps the encoded conversation into a latent distribution), and a decoder Ψ. The stochastic encoder $(\Phi, \mathcal{G}_0)$ comprises an encoder Φ that maps tokenized conversation histories $\mathbf{X}$ to a latent space $\mathcal{Z} \subseteq \mathbb{R}^d$, i.e., $z = \Phi(\mathbf{X}) \in \mathcal{Z}$, which is then used to construct a parameterized d-

[2]Note that we use Y as the next utterance in the dataset and $\hat{Y}$ as the one predicted by the LM.

dimensional Gaussian distribution $\mathcal{G}_0(z'|z) = \mathcal{N}\big(\mu_0(z), \sigma_0^2(z)\mathbf{I}_{d\times d}\big)$ over $\mathbb{R}^d$. The decoder predicts the next utterance $\widehat{Y}_0$ (token-by-token) conditioned on the point z' sampled from the latent distribution, i.e., $\Psi(\widehat{Y}_0|z')$[3], $z' \sim \mathcal{G}_0(\cdot|z)$. We denote by $\texttt{LM}_0(Y|\mathbf{X}) := \mathbb{E}_{z'\sim\mathcal{G}_0(\cdot|z), z=\Phi(\mathbf{X})}[\Psi(Y|z')]$, the *primitive* and learn it using a loss function that in addition to predicting the next utterance accurately, encourages diversity and generalization in the learned latent space $\mathcal{Z}$ (see Eq. 1 and Fig. 2). As we will explain in Section 4, our loss function is inspired by those in prior work, and more specifically by the one in OPAL (Ajay et al., 2020), i.e., an unsupervised learning method for discovering primitive skills in trajectories that are used by some downstream RL tasks.

Step 2: Expert Construction. Given the latent space $\mathcal{Z}$, encoder $(\Phi, \mathcal{G}_0)$, and decoder Ψ learned in Step 1, we now learn m latent distributions $\{\mathcal{G}_i\}_{i=1}^m$, each defined as $\mathcal{G}_i(z'|z) = \mathcal{N}\big(\mu_i(z), \sigma_i^2(z)\mathbf{I}_{d\times d}\big)$. Intuitively, each $\mathcal{G}_i$ corresponds to an attribute, e.g., an intent or a personality (in case of a chatbot) and generates samples in specific parts of the latent space $\mathcal{Z}$. This results in having m LMs, $\{\texttt{LM}_i\}_{i=1}^m$, $\texttt{LM}_i = (\Phi, \mathcal{G}_i, \Psi)$, each of them corresponds to a *specialized* version of the original LM, $\texttt{LM}_0$, and serves as an *expert* in our MoE-LM. Upon receiving a conversation history $\mathbf{X}$, each expert $\texttt{LM}_i$ generates a candidate (or more) for the next utterance $\widehat{Y}_i$ in certain parts of the language space that are compatible with its attribute (personality). As we will explain in Section 5, each $\mathcal{G}_i$ is learned using a loss function that encourages its corresponding LM, $\texttt{LM}_i$, to generate utterances consistent with its attribute (see Eq. 2).

Step 3: Dialogue Manager (DM). The *dialogue manager*, denoted by μ, takes as input the encoded conversation history $z = \Phi(\mathbf{X})$ and the candidate action utterances generated by the experts $\{\widehat{Y}_i\}_{i=0}^m$, and selects one of them as the action for the bot to execute, i.e., $\widehat{Y} \sim \mu(\cdot \mid z, \{\widehat{Y}_i\}_{i=0}^m)$. We will describe how DM is trained using reinforcement learning (RL) in Section 6.

4 Primitive Discovery in MoE-LM

Motivated by literature in the reinforcement and imitation learning fields (Ajay et al., 2020), we propose to learn the primitive LM, $\texttt{LM}_0$, in our MoE-LM by solving the following KL-constrained optimization problem that aims at capturing diverse semantics:

$$\min_{(\Phi,\mathcal{G}_0,\Psi),\rho} \widehat{\mathbb{E}}_{z'\sim\rho(\cdot|z,Y), z=\Phi(\mathbf{X})}\big[-\log\Psi(Y|z')\big], \quad \text{s.t. } \widehat{\mathbb{E}}_{z=\Phi(\mathbf{X})}\big[\mathrm{KL}\big(\rho(z'|z,Y)\;||\;\mathcal{G}_0(z'|z)\big)\big] \le \epsilon_{\mathrm{KL}}, \tag{1}$$

where $\widehat{\mathbb{E}}$ is the empirical expectation over $(\mathbf{X}, Y)$ in the dataset $\mathcal{D}$, ρ is a distribution over the latent space conditioned on the encoded conversation history z and the target utterance Y, and ϵ_{KL} is a positive real-valued threshold. Using (1), we learn $\texttt{LM}_0 = (\Phi, \mathcal{G}_0, \Psi)$ by maximizing the log-likelihood, while enforcing consistency between the latent variable z' predicted by $\mathcal{G}_0(\cdot|z)$ and $\rho(\cdot|z, Y)$ via the KL constraint. The distribution $\rho(\cdot|z, Y)$ is a Gaussian $\mathcal{N}\big(\mu_\rho(z, \Phi_\rho(Y)), \sigma_\rho^2(z, \Phi_\rho(Y))\mathbf{I}_{d\times d}\big)$ in which Φ_ρ is a pre-trained encoder for the target utterance Y, and mean $\mu_\rho(\cdot,\cdot)$ and variance $\sigma_\rho^2(\cdot,\cdot)$ are trainable models. One reason for using a separate encoder Φ_ρ for the target utterance Y is to avoid overfitting Φ (i.e., to avoid having back-propagation gradient of Φ with Y as input).

Connection to VAE-like objectives In practice, we implement the KL constraint in (1) as a penalty weighted by an appropriately chosen coefficient. Thus, one may interpret the objective in (1) as a variation of β-VAE (Burgess et al., 2018). Due to the connection to VAEs, one may draw similarities between our method and existing dialogue approaches such as VHRED (Serban et al., 2017b) and VHCR (Park et al., 2018). However, we emphasize that there are key differences, and these may be explained by first understanding how the objective in (1) encourages *diversity*, which is key to good primitive learning. Namely, it is important that primitive discovery learns an encoder-decoder Φ, Ψ which can be modulated by the choice of z; i.e., changing z' while fixing $\mathbf{X}$ should lead to different distributions over generated utterances. The objective in (1) encourages this diversity by conditioning the latent variable z' on both the target utterance Y and $z = \Phi(\mathbf{X})$, i.e., $z' \sim \rho(\cdot|z, Y)$. In contrast, the KL constraint is used to make sure that the stochastic encoder $\mathcal{G}_0(\cdot|z)$ of our primitive LM is not too varied for different Y, and thus has a *limiting* effect on diversity. For example, in the extreme when $\epsilon_{\mathrm{KL}} = 0$ (or, $\beta \to \infty$ when used as a regularizer) there will be no specialization of the latent space for different Y. Although $\beta \to \infty$ is an extreme case, degenerate behavior can also happen when $\beta = 1$, i.e., in the traditional variational loss used by VHRED and VHCR. Specifically, it is well-known that the traditional VAE loss is an upper bound on the negative log-likelihood of the data under a stochastic encoder-decoder parameterization. Thus if the data can be modeled by a single

[3]In practice, we can use both latent states as the input to the decoder model $\Psi(\widehat{Y}_0|z', z)$.

LM then a VAE-optimal decoder Ψ can simply ignore $\mathcal{G}_0$, leading to a degenerated latent space as observed in previous work (Park et al., 2018). This is precisely the reason that, in our approach, we weaken the KL constraint ($\epsilon_{\text{KL}} \gg 0$ or, equivalently, $\beta \ll 1$). This enables our approach to more reliably guarantee that a unique z' represents each distinct conversation pair $(\mathbf{X}, Y)$, thus capturing diverse semantic modalities and enabling easier downstream specialization.

In the mathematical results below, we formalize the claim above, namely, that the log-likelihood objective in (1) leads to a learned Φ, Ψ that can easily recover any arbitrary desired LM by specializing the latent space $\mathcal{G}$. We begin with a definition that characterizes the coverage of an arbitrary LM on the conditional conversation data distribution $P_{\mathcal{D}}(Y|\mathbf{X})$.

Definition 1. $\texttt{LM}_{\mathcal{D},\xi}$ is a ξ-common LM of data $\mathcal{D}$ if $\mathbb{E}_{\mathcal{D}}[\text{TV}(\texttt{LM}_{\mathcal{D},\xi}(Y|\mathbf{X})||P_{\mathcal{D}}(Y|\mathbf{X})))] \leq \xi$.

Leveraging Theorem 4.1 in Ajay et al. (2020), we now present the theoretical result characterizing the representational power of our primitive encoder-decoder pair (Φ, Ψ) on data $\mathcal{D}$.

Lemma 1. Let (Φ, ρ, Ψ) be the solution to (1) with $\widehat{\mathbb{E}}_{z'\sim\rho(\cdot|z,Y),z=\Phi(\mathbf{X})}[-\log \Psi(Y|z')] = \epsilon$. Then there exists $\texttt{LM} := (\Phi, \mathcal{G}, \Psi)$ such that $\mathbb{E}_{\mathcal{D}}[\text{TV}(\texttt{LM}_{\mathcal{D},\xi}(Y|\mathbf{X})||\texttt{LM}(Y|\mathbf{X}))] \leq \xi + \sqrt{\frac{1}{2}(\epsilon + \mathcal{H})}$, where $\mathcal{G}(z'|z) = \mathbb{E}_{Y\sim\mathcal{D}}[\rho(z'|z, Y)]$, and $\mathcal{H} = \mathbb{E}_{\mathcal{D}}[\log P_{\mathcal{D}}(Y|\mathbf{X})]$ is a constant depending on $\mathcal{D}$.

The result above shows that, as long as $\texttt{LM}_{\mathcal{D},\xi}$ is ξ-common in $\mathcal{D}$, then there exists a specialization of the latent space $\mathcal{G}$ that, when paired with Φ, Ψ, can approximately recover $\texttt{LM}_{\mathcal{D},\xi}$. The quality of the approximation is a function of ϵ — how well the objective in (1) was optimized — and ξ. In practice, to construct the primitive by replacing $\mathcal{G}$ with $\mathcal{G}_0$, i.e., $\texttt{LM}_0 = (\Phi, \mathcal{G}_0, \Psi)$, because $\mathcal{G}_0(z'|z)$ can be viewed as an distillation of $\rho(z'|z, Y)$. This theoretical result also motivates the next section, where we explain our algorithm's "Step 2: Expert Construction". Specifically, we show how to use the trained encoder-decoder pair Φ, Ψ to learn a spectrum of different specialized experts parameterized by different latent distributions $\mathcal{G}_i$.

5 Expert Construction with Plug-and-play Language Models

To complete the MoE framework one needs to systematically create a gamut of different experts $\texttt{LM}_i, \forall i \in \{1, \dots, m\}$, with each generating candidate utterances of different intents. By viewing each expert as a distribution of particular behaviors in conversation data $\mathcal{D}$, we leverage the results of Section 4 and Lemma 1 and adopt a universal encoder-decoder (Φ, Ψ) among all the experts. Therefore, each expert i is only parameterized by an arbitrary d-dimensional latent distribution (e.g., Gaussian), and it samples certain regions of the latent space $\mathcal{Z}$. Following the terminology from Dathathri et al. (2019), these experts can all be catagorized as *plug-and-play language models (PPLMs)*. Creating experts is handy because it only requires learning new latent distributions, while switching between experts amounts to sampling a different distribution.

Denote by $\ell_i(\mathbf{X}, Y) \in \mathbb{R}$ a real-valued label that *characterizes* the intent of expert $i \in \{1, \dots, m\}$, e.g., determined by an off-the-shelf sentiment classifier. We train the latent distribution $\mathcal{G}_i(z)$ of expert i by solving the optimization problem

$$\min_{\mathcal{G}_i} \widehat{\mathbb{E}}_{z'\sim\mathcal{G}_i(\cdot|z),z=\Phi(\mathbf{X}),Y\sim\Psi(\cdot|z')}[-\ell_i(\mathbf{X},Y)]. \tag{2}$$

Unlike the weighted maximum likelihood approach considered in Dathathri et al. (2019), which assigns weight ℓ_i to training samples that correspond to expert i, we propose to learn each expert via *reward-maximization* and treat ℓ_i as a reward signal w.r.t. expert i to be maximized. Interestingly, this approach is also linked to reinforcement learning (RL), in which both the "state" and "action" spaces are the latent space $\mathcal{Z}$, and the "policy" is the latent distribution $\mathcal{G}_i$. The main benefit of our approach is that it does not require the target utterance Y from data $\mathcal{D}$ and is thus less vulnerable to data-imbalance issues in $\mathcal{D}$ on certain intents. Notice from (2) that the reward-maximization problem is myopic, i.e., the above RL problem has a discounting factor of 0. The main motivation is that, unlike dialogue management that is a sequential decision-making problem, here we want each expert to possess particular behaviors, and this can readily be done via greedy maximization. Long-term dialogue optimization will be handled by the dialogue manager rather than these experts.

For example in the case of a Gaussian $\mathcal{G}_i$, we use the standard REINFORCE (Sutton et al., 1999) algorithm to learn the model parameters (μ_i, σ_i^2) of $\mathcal{G}_i$ according to

$$\{\mu_i, \sigma_i\} \leftarrow \{\mu_i, \sigma_i\} + \alpha \cdot \mathbb{E}_{z'\sim\mathcal{G}_i(\cdot|z),Y\sim\Psi(\cdot|z')}[\ell_i(\mathbf{X}, Y) \cdot \nabla_{\{\mu_i,\sigma_i\}} \log \mathbb{P}_{\mathcal{G}_i}(z'|z)],\ i \in \{1, \dots, m\},$$

where $\alpha > 0$ is the learning rate. To reduce the variance of these estimates, we also adopt the baseline reduction technique (Greensmith et al., 2004) in policy gradient, which replaces $\ell_i(\mathbf{X}, Y)$

with $\overline{\ell}_i(\mathbf{X}, Y) := \ell_i(\mathbf{X}, Y) - \mathbb{E}_{Y\sim\Psi(\cdot|\Phi(\mathbf{X}))}[\ell_i(\mathbf{X}, Y)]$. Following arguments from Lemma 1 and Lemma 4.0.1 in Ajay et al. (2020), in the following we quantify the sub-optimality of expert $\texttt{LM}_i$.

Corollary 1. Denote the i-th reward-maximizing objective as $\mathcal{L}_i(\texttt{LM}) := \widehat{\mathbb{E}}_{Y\sim\texttt{LM}(\cdot|\mathbf{X})}[\ell_i(\mathbf{X},Y)]$. Suppose an optimal LM for this objective $\texttt{LM}_{i,\xi} \in \arg\max_{\texttt{LM}} \mathcal{L}_i(\texttt{LM})$ is ξ-common in $\mathcal{D}$. Moreover, let $\mathcal{G}_i^\star$ be in the $\arg\min$ of (2). Then with expert $\texttt{LM}_i = (\Phi, \mathcal{G}_i^\star, \Psi)$ and $(\epsilon, \mathcal{H})$ from Lemma 1, we have $|\mathcal{L}_i(\texttt{LM}_i) - \mathcal{L}_i(\texttt{LM}_{i,\xi})| \le 2\|\ell_i\|_\infty \cdot (\xi + \sqrt{\frac{1}{2}(\epsilon + \mathcal{H})})$.

While it may be obvious that optimizing $\mathcal{G}_i$ w.r.t. (2) encourages expert $\texttt{LM}_i$ to capture the behaviors encouraged by ℓ_i, this corollary has two further implications: (i) Since the sub-optimality of $\texttt{LM}_i$ compared to the oracle $\texttt{LM}_{i,\xi}$ is bounded by the quantity ϵ defined in Lemma 1, it justifies using the primitive (Ψ, Φ), which optimizes ϵ, for expert construction; (ii) Sub-optimality further depends on ξ, quantifying how well $\texttt{LM}_{i,\xi}$ is represented in the original dataset $\mathcal{D}$.

6 Reinforcement Learning for MoE-LM Dialogue Manager

We now describe the dialogue manager (DM) of our MoE-LM and propose RL algorithms to train it. As mentioned in Section 3, the DM is a policy μ that takes the encoded conversation history $z = \Phi(\mathbf{X})$ and the $m + 1$ candidate action utterances generated by the experts $\{\widehat{Y}_i\}_{i=0}^m$,[4] and stochastically selects one of them to execute, i.e., $\widehat{Y} \sim \mu(\cdot \mid z, \{\widehat{Y}_i\}_{i=0}^m)$. Note that each expert $i \in \{0, \dots, m\}$ is a LM, $\texttt{LM}_i$, that acts as a policy $\pi_i(\cdot|\mathbf{X})$ and maps each conversation history $\mathcal{X}$ to an utterance $\widehat{Y}_i$. With this architecture we address the large size of state and action spaces in the original MDP that grows exponentially with the size of the vocabulary. As described in Section 2, the state and action spaces of the original MDP are the tokenized conversation history and the tokenized language space, respectively, while here the DM should choose among $m + 1$ actions (which is a much smaller and finite action space) given the latent space $\mathcal{Z}$ (which is a continuous state space) of encoded conversations. It is important to note that our MoE-LM is different than other hierarchical architectures (Kulkarni et al., 2016) in which the decision at the high-level is to choose a low-level controller only based on the current state of the system. In MoE-LM, the DM observes both the current state and the actions suggested by the experts and then chooses one among them.

We consider two RL settings to solve this specialized MDP. The first one is offline RL, in which the policy must be learned from the collected conversations $\mathcal{D}$ without further (online) interactions. Offline RL requires optimizing a policy, while minimizing the deviation from the behavior policy to avoid errors due to data co-variate shift. Among numerous offline RL algorithms (Kumar et al., 2020; Carta et al., 2021; Jaques et al., 2019), one effective algorithm to learn the DM policy μ is IQL (Kostrikov et al., 2021). Given conversation data $\mathcal{D}$, IQL first computes the critic functions $(Q_{\theta^*}(z, a), V_{\phi^*}(z))$ via solving $\min_{\theta,\phi} \mathbb{E}_{(z,a,r,z_+)\in\mathcal{D}}[(r + \gamma V_\phi(z_+) - Q_\theta(z, a))^2] + \lambda \cdot \mathbb{E}_{(z,a)\in\mathcal{D}}[L_2^\tau(Q_\theta(z, a) - V_\phi(z))]$, where z is the encoded conversation, a is the bot utterance, z_+ is the next encoded conversation, r is the conversation-level reward, $\lambda > 0$ is a tunable weight, and L_2^τ is the expectile regression operator (Koenker and Hallock, 2001) of estimating the top-τ expectile statistics of the Q-function random variable (approximated by the value function V_ϕ), and then extracts the DM policy μ via greedification over the finite set of MoE candidate utterances: $\mu(a \mid z, \{\widehat{Y}_i\}_{i=0}^m) = \arg\max_{a\in\{\widehat{Y}_i\}_{i=0}^m} Q_{\theta^*}(z, a)$. Intuitively, IQL leverages the generalization capacity of critic functions to estimate the value of the best action without directly querying the values w.r.t. unseen actions. This makes it less conservative than most offline RL methods that either constrain the policy's actions to be in-distribution or solve a behavior-regularized policy optimization problem.

The second RL setting for learning the DM policy μ is via model-based RL (MBRL) (Shah et al., 2018; Wei et al., 2018). While this paradigm can be applied to any online/offline RL algorithms we demonstrate it with the simple DQN (Mnih et al., 2013). Here we first learn a user utterance model $P_{\text{user}}(X_+|\mathbf{X}, a) := \mathbb{E}_{z=\Phi_{\text{user}}([\mathbf{X},a])}[\Psi_{\text{user}}(X_+|z)]$ via maximum likelihood, then generate data $\mathcal{D}_{\text{MB}}$, whose next-state $\widehat{s}_+$ encodes the next conversation generated from roll-outs and the corresponding candidate actions, solve the Bellman error minimization: $\min_\theta \sum_{(s,a,r,\widehat{s}_+)\in\mathcal{D}_{\text{MB}}} (\overline{r} + \gamma Q_{\theta^{\text{target}}}(\widehat{s}_+, \arg\max_{a_+\in\{0,\dots,m\}} Q_\theta(\widehat{s}_+, a_+)) - Q_\theta(s, a))^2$, and obtain the DM policy μ via the aforementioned greedification step. The benefit of MBRL over offline RL is two-fold. First, one can easily obtain a *high-fidelity* user utterance model (Peng et al., 2020) by simply fine-tuning a large LM (e.g., GPT-3 (Floridi and Chiriatti, 2020)). Second, with sufficient

[4]For simplicity, we assume that each expert generates only a single candidate utterance at each step. It would be straightforward to extend this to multiple (and even a varying number of) candidate utterances.

dialogue roll-outs that captures many different scenarios, MBRL generally can be more data-efficient and less prone to distributional shifts than offline RL.

7 EXPERIMENTS

We evaluate our MoE-approach on two open-domain benchmarks that are common within the RL-based dialogue management literature (Jaques et al., 2019). The first one is the Cornell Movie corpus (Danescu-Niculescu-Mizil and Lee, 2011), which consists of conversations between speakers in different movie lines and has a median conversation length of 3 utterances. The second is the Reddit Casual (Ghandeharioun et al., 2019) conversations, which is a subset of the Reddit corpus that only contains casual conversations on various topics of at least 3 turns and a median of 7 utterances.

We conduct several experiments to test the efficacy of different parts in the MoE-LM, namely (i) the predictive power and diversity of the primitive, (ii) the quality of experts, and (iii) the overall DM performance. For each metric, we report mean $\pm$ standard error over 100 conversations sampled from the evaluation set. We also ran an ablation study on 4 transformer-based MoE-LMs, namely MoE-1, MoE-2, MoE-3- MoE-4, to understand how performance is affected by different model architectures, language encoders, and latent generators. MoE-1 and MoE-2 use a simpler architecture, while MoE-3 and MoE-4 use the same encoder architecture as BERT (Devlin et al., 2018). MoE-1 uses much smaller latent distribution models $\{\mathcal{G}_i\}$ than MoE-2; MoE-3 uses the pre-trained BERT encoder, while MoE-4 trains that from scratch. Details of these models can be found in Appendix B.3.

EXP 1: Comparing Primitive Models We compare the quality of latent representations learned by the 4 MoE-LMs (via optimizing Eq. 1) and 2 baselines (standard Transformer (Vaswani et al., 2017) and VHRED [5] (Serban et al., 2017b)). To assess their quality, for each test conversation we generated 25 utterances and reported the following 3 metrics: (i) **Diversity**: The $1-$ Sparsity (Hurley and Rickard, 2009) of the singular values of the embedded utterances, i.e., Diversity$(\{\hat{Y}_i\}) := 1 - \sqrt{d - \|\text{SVD}\|_1/\|\text{SVD}\|_2}/\sqrt{d-1} \in [0, 1]$, where $\text{SVD} := \text{SVD}(\{\Phi_{\text{SE}}(\hat{Y}_i\}_{i=1}^{25})$, and Φ_{SE} is a pre-trained sentence encoder (e.g., a USE (Cer et al., 2018)); (ii) **Dist-$\{1, 2, 3\}$** (Li et al., 2015): Ratio of unique $\{1, 2, 3\}$-gram in the generated utterances; (iii) **Perplexity**: The perplexity score of the utterance w.r.t. a GPT-2 LM, which is more correlated to human's judgement on semantic fluency (Pang et al., 2020). These metrics measure both accuracy and semantic diversity. Qualitatively, we also measure fluency and diversity of LMs using human ratings (see Appendix B.8 for details). The results of the above experiments are reported in Table 1 and 6 (Appendix A.1), and sample utterances generated by these LMs can be found in Appendix A.3. Human evaluation on diversity/fluency of different LMs are given in Table 4. In comparisons with the baselines (Transformer and VHRED), generally (i) transformer-based LMs out-perform VHRED due to their attention mechanism that explicitly encodes sequential semantic information, and (ii) the MoE-LMs achieve way better diversity without sacrificing much on accuracy (i.e., the perplexity scores are still quite low). Qualitatively, the sample utterances generated the Transformer are closer to the targets than that by MoE-2 and MoE-4, likely because Transformer tends to memorize the corpus (Kharitonov et al., 2021). Contrarily, MoE-LMs generate utterances that have similar contexts with targets but paraphrased or similar structures but different contexts, demonstrating their generalizability. Human evaluations also show that MoE-2 and MoE-4 generate more diverse utterances while retaining sufficient semantic fluency.

Among different MoE-LMs, MoE-2 and MoE-4 have the best performances, particularly MoE-4 has better diversity while MoE-2 has lower perplexity. This corroborates with our hypotheses that (i) jointly training the encoder and decoder with Eq. 1 seems necessary to encourage semantic diversity (as opposed to using a pre-trained BERT encoder, which maximizes likelihood), (ii) sufficient representation power is necessary for $\mathcal{G}_0$ to match the posterior distribution ρ in order to capture different semantics in $\mathcal{D}$. In Fig. 2a and 2b, we visualize the latent space of 400 conversation data samples for both Transformer and MoE-4. The latent states of MoE-4 are much more dispersed across the embedding space, implying that most conversations get encoded uniquely. In contrast, the latent space of Transformer has many clusters, suggesting it is more prone to generating similar utterances even with different input conversation and is thus less generalizable.

EXP 2: Quality of Experts We compare the performance of experts learned by the 4 MoE-LMs (where experts are separately trained by optimizing Eq. 2) and 2 baselines (WD (Holtzman et al., 2018) and PPLM (Dathathri et al., 2019)). To study the sub-optimality gap in Corollary 1, we also include the performance of Transformer-based expert end-to-end LMs that are individually optimized with REINFORCE (Li et al., 2016), using the expert labels $\{\ell_i\}$ as rewards. Inspired by Ghandeharioun et al. (2019) on how bot behaviors are characterized, we use the following label functions to define the

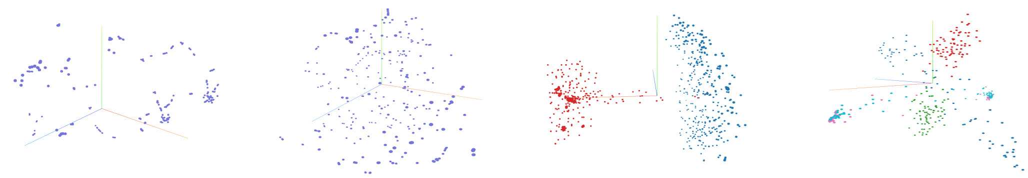
[5] The VHRED model implementation is identical to that in Jaques et al. (2019) to ensure fair comparisons.

Method	Diversity	Dist-1	Dist-2	Dist-3	Perplexity
MoE-1	0.069 ± 0.03	0.27	0.66	0.75	27.43 ± 18.49
MoE-2	0.14 ± 0.05	0.35	0.77	0.90	38.81 ± 17.34
MoE-3	0.089 ± 0.04	0.29	0.75	0.90	41.35 ± 26.68
MoE-4	$\mathbf{0.16 \pm 0.04}$	**0.38**	**0.80**	**0.95**	50.17 ± 28.11
Trans.	0.087 ± 0.03	0.26	0.65	0.85	$\mathbf{19.23 \pm 15.46}$
VHRED	0.09 ± 0.04	0.35	0.70	0.79	79.77 ± 39.61

Table 1: Accuracy (Perplexity) and Diversity of Language Primitive Experts Trained with Reddit.

Method	User Tot. Sent.	User Sent. Trans.	Perplexity
MoE-4 IQL	$\mathbf{4.55 \pm 0.38}$	$\mathbf{2.88 \pm 0.35}$	45.53 ± 26.71
MoE-4 Ens-Q	4.14 ± 0.41	2.36 ± 0.31	43.77 ± 28.24
MoE-4 KLC	3.94 ± 0.25	2.21 ± 0.24	$\mathbf{38.35 \pm 16.88}$
VHRL	3.85 ± 0.28	2.19 ± 0.28	55.81 ± 24.21
VHRL-KLC	3.95 ± 0.19	2.16 ± 0.33	64.05 ± 36.98
VHRL-SAC	3.93 ± 0.28	2.19 ± 0.32	62.06 ± 40.43

Table 2: Performance (w.r.t. User Satisfaction in Conversation) of MBRL-based DM Trained with Reddit.



(a) Transformer Primitive, TSNE (b) MoE-4 Primitive, TSNE (c) MoE-4 Sentiment, PCA (d) MoE-4 Emotion, PCA

Figure 2: Latent Space Visualizations. Figures (a) and (b) Compares Two Primitive Representations. Figures (c) and (d) Illustrates How Experts (of Different Sentiments and Emotions) are Represented by Latent Clusters.

intents of experts: (i) $\ell_{\text{pos-sent}}(Y)$, $\ell_{\text{neg-sent}}(Y)$, $\ell_{\text{joy}}(Y)$, $\ell_{\text{optimism}}(Y)$, $\ell_{\text{anger}}(Y)$, $\ell_{\text{sadness}}(Y)$ quantify 6 different sentiment tones and are constructed by a RoBERTa-based sentiment detector (Liao et al., 2021) that predicts whether an utterance is of positive or negative sentiment, and whether it falls into any of the 4 more-refined emotions: {joy, optimisim, sadness, anger}; (ii) $\ell_{\text{sent-coh}}(\mathbf{X}, Y)$ measures *empathy*, i.e., bot's sentiment coherence with user's; (iii) $\ell_{\text{question}}(Y)$ outputs 1 when a bot question is detected and 0 otherwise; (iv) $\ell_{\text{exp}}(\mathbf{X}, Y)$ quantifies *exploration*, i.e., the tendency to avoid repetitive contexts (see Appendix B.7 for details). Qualitatively, we also measure fluency and expert skills of LMs using human ratings (see Appendix B.8 for details). The results of the above experiments are reported in Table 3 and 8 (Appendix A.1), with sample utterances reported in Appendix A.4 to A.10. Results on human evaluation of different LMs w.r.t. fluency and different expert skills are given in Table 5. Compared with the baseline LMs, generally the experts created under the MoE-LM framework, especially under MoE-2 and MoE-4, better capture all different language intents (where WD and PPLM appear to capture negative sentiments and emotions much more effectively than behaviors), demonstrating the efficacy of our approach which constructs specialized experts on a diverse language space via reward maximization (instead of weighted MLE). Human evaluations also show that MoE-4 is most effective in generating semantically fluent utterances that possess a wide range of expert characteristics.

Similar to the ablation study in EXP 1, all the experts associated with MoE-2 and MoE-4 are also among the best ones in capturing language intents. Interestingly, with the Reddit data the experts in MoE-4 perform the best, while with much less data (Cornell) the best experts are built upon the simpler MoE-2 architecture. We conjecture this difference is due to over-fitting issues faced by the larger LMs (MoE-4) when there is insufficient data for expert fine-tuning. In Fig. 2c and 2d we visualize the latent space of the sentiment-based experts in MoE-4, each with 400 conversation data samples. Notice that the sentiment experts' latent distributions are clearly separated (because positive and negative sentiments have opposite behaviors), while the emotion expert's latent distribution have more gradual separations and even some overlaps (because e.g., joy versus optimism are quite similar, while joy versus anger are quite different). This validates our MoE-LM represents different behaviors in separate regions of the latent space and justifies our structural prior of modeling each expert as a specialized version of the primitive LM, whose latent distribution focuses on particular regions.

EXP 3: MoE-RL Against DialoGPT Simulated Users We compare the dialogue management performance of MoE-LM, for which their MoE-based DMs μ are trained with different methods: (i) IQL (Kostrikov et al., 2021), (ii) Ensemble DQN (Carta et al., 2021), (iii) KL-control (Jaques et al., 2019), with 3 standard RL-based DM baselines using the VHRL architecture (Saleh et al., 2020): (i) REINFORCE (Li et al., 2016), (ii) KL-control, (iii) SAC (Haarnoja et al., 2018). According to the results on expert quality in EXP2, we pick the MoE-2 and MoE-4 frameworks for the Cornell and Reddit tasks respectively. For systematic evaluation, we perform the experiment by having these RL agents interact with a DialoGPT (Zhang et al., 2019) simulated user environment for a maximum of 5 turns. The DM task is to maximize total user satisfaction in the conversation level, which is measured by both (i) user's overall sentiment, and (ii) user's sentiment transition. To construct an immediate reward that serves as a surrogate for user satisfaction, we set $r(\mathbf{X}, a, X_+) = \lambda_1 \ell_{\text{sent}}(X_+) + \lambda_2(\ell_{\text{sent}}(X_+) - \frac{1-\gamma}{1-\gamma^L} \sum_{l=0}^{L-1} \gamma^l \ell_{\text{sent}}(X_l))$, where the linear combi-

Method	Question	Exploration	Positive Sent.	Negative Sent.	Sent. Coherence	Joy	Optimism	Anger	Sadness
MoE-1	0.65 ± 0.20	0.45 ± 0.17	1.13 ± 0.21	0.35 ± 0.19	0.50 ± 0.38	0.96 ± 0.26	-0.21 ± 0.56	0.54 ± 0.58	0.99 ± 0.83
MoE-2	0.95 ± 0.27	0.47 ± 0.21	3.29 ± 0.33	1.42 ± 0.38	0.51 ± 0.40	1.99 ± 0.38	1.25 ± 0.43	$\mathbf{1.48 \pm 0.39}$	$\mathbf{2.01 \pm 0.46}$
MoE-3	0.41 ± 0.35	0.50 ± 0.24	1.23 ± 0.78	0.99 ± 0.48	$\mathbf{0.66 \pm 0.35}$	1.02 ± 0.29	0.49 ± 0.51	0.53 ± 0.49	1.10 ± 0.48
MoE-4	$\mathbf{0.96 \pm 0.37}$	$\mathbf{0.51 \pm 0.31}$	$\mathbf{3.41 \pm 0.55}$	$\mathbf{1.80 \pm 0.34}$	0.52 ± 0.31	$\mathbf{2.05 \pm 0.55}$	$\mathbf{1.57 \pm 0.44}$	1.42 ± 0.42	1.97 ± 0.36
WD	0.05 ± 0.03	0.15 ± 0.37	-0.50 ± 0.74	1.01 ± 0.48	0.51 ± 0.20	-0.51 ± 0.39	-0.84 ± 0.76	1.00 ± 0.44	1.27 ± 0.67
PPLM	0.20 ± 0.25	0.48 ± 0.28	0.44 ± 0.41	0.69 ± 0.22	0.53 ± 0.31	0.31 ± 0.29	0.40 ± 0.55	0.71 ± 0.46	0.98 ± 0.59
Trans. RL*	0.99 ± 0.23	0.54 ± 0.18	3.53 ± 1.64	1.89 ± 1.20	0.72 ± 0.30	2.88 ± 0.96	1.80 ± 0.59	1.62 ± 0.75	2.35 ± 0.62

Table 3: Quality of Each Expert PPLM Trained on Reddit Casual dataset w.r.t. its Trained Label. Trans. RL Corresponds to Individually Optimized LMs Using Expert Labels $\{\ell_i\}$ as Rewards.

Method	Avg. Fluency	Diversity
MoE-1	0.72 ± 0.02	0.51 ± 0.02
MoE-2	$\mathbf{0.75 \pm 0.02}$	0.54 ± 0.02
MoE-3	0.72 ± 0.02	0.42 ± 0.03
MoE-4	0.65 ± 0.03	$\mathbf{0.69 \pm 0.02}$
Trans.	0.70 ± 0.02	0.47 ± 0.02

Table 4: Phase 1 Raters Evaluation

Method	Avg. Fluency	S_{Question}	$S_{\text{Pos. Sent}}$	$S_{\text{Neg. Sent}}$	S_{Joy}	S_{Anger}
MoE-3	0.76 ± 0.04	0.27 ± 0.05	$\mathbf{0.64 \pm 0.04}$	0.26 ± 0.05	0.41 ± 0.04	0.33 ± 0.05
MoE-4	$\mathbf{0.82 \pm 0.02}$	$\mathbf{0.74 \pm 0.04}$	0.46 ± 0.04	$\mathbf{0.44 \pm 0.03}$	$\mathbf{0.51 \pm 0.04}$	$\mathbf{0.43 \pm 0.03}$
Trans.	0.78 ± 0.03	0.12 ± 0.03	0.29 ± 0.04	0.19 ± 0.04	0.38 ± 0.04	0.33 ± 0.05

Table 5: Phase 2 Raters Evaluation (Reddit Casual Models).

nation weights $(\lambda_1, \lambda_2) = (0.75, 0.25)$ correlate with Ghandeharioun et al. (2019), and $\ell_{\text{sent}}(X)$ is the same RoBerTa-based sentiment labeler as in EXP2, which assigns a score from $[-1, 1]$ that is proportional to the positive sentiment and inversely proportional to the negative sentiment prediction probabilities. To ensure the baseline RL DM methods can also possess certain bot-level features, e.g., question, positive sentiment, etc., besides the above RL reward for user satisfaction we also optimize a linear combination of bot-based rewards when training the baseline models, see Appendix B of Saleh et al. (2020) for more details. Since the DM problem lasts at most 5 turns, we use this as the effective horizon and set $\gamma = 1 - 1/5 = 0.8$.

The results of the above experiments (performed in both offline RL and MBRL settings) are reported in Table 2, Table 7 (Appendix A.1), Table 9 and 10 (Appendix A.2), with sample utterances reported in Appendix A.11. Our experiments show that MoE-LMs outperform most baselines on DM performance. We attribute this finding to three factors: (i) MoE-LM restricts the action space into a smaller set of candidate utterances generated by experts (whose qualities are validated in EXP2), the corresponding RL problem then becomes simpler and requires less data (especially in Cornell) to solve; (ii) Unlike the baseline RL methods, which need to optimize both bot-and-user signals, the MoE DM agents focus on optimizing the user satisfaction goal and are therefore more effective; (iii) Among different RL settings, MBRL, which first learns a user utterance model (the model uses the same encoder from the primitive LM and learns a separate decoder for user-utterance prediction) then does DM, performs much better than offline RL methods that moderately improve upon the primitive LM (behavior policy). IQL-based dialogue managers are among the best across different settings potentially because IQL is more robust to co-variate shifts than standard RL methods, e.g., Ens-Q, SAC, and yet it is less conservative than the behavior-regularized algorithms, e.g., KLC. Interestingly, our MoE-LMs also have lower (better) perplexity scores than other methods. This may be due to the fact that MoE-LM uses pre-trained encoder and decoder from the primitive LM, which are optimized for generalization and accuracy, while other RL methods may distort their language representations to create utterances that maximize reward but become less natural.

8 Concluding Remarks

We developed a mixture-of-expert (MoE) approach for RL-based dialogue management (DM). Our MoE language model (MoE-LM) comprises of three main components: (i) a LM that can generate diverse semantics for conversation histories, (ii) a number of specialized LMs (or experts) that can produce utterances corresponding to a particular attribute or intent, and (iii) a RL-based DM that performs dialogue planning with the utterances generated by the experts. To understand how well our MoE approach generates diverse and sensible utterances, and solves DM problems, we evaluated it using two open-domain dialogue tasks and compared it with SOTA baselines. Our results showed that MoE-LM (i) improves diversity of text generation, (ii) can generate utterances with specific intents, and (iii) yields better overall performance. We consider our work as a step forward in creating steerable LMs that possess different intents and in training RL-based DMs that can carry on rich conversations. Future work includes improving the language representation with information-theoretic approaches, fine-tuning the experts based on the DM objective, extending the RL agent to track users' behaviors (via abstract belief states) and plan upon them, preventing RL dialogue agents from generating harmful behaviors (via enforcing safety constraints), and evaluating our MoE-LM on more realistic problems, such as information retrieval, recommendation, and negotiation.

References

A. Ajay, A. Kumar, P. Agrawal, S. Levine, and O. Nachum. Opal: Offline primitive discovery for accelerating offline reinforcement learning. *arXiv preprint arXiv:2010.13611*, 2020.

K. Asadi and J. Williams. Sample-efficient deep reinforcement learning for dialog control. *arXiv preprint arXiv:1612.06000*, 2016.

C. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner. Understanding disentangling in β-vae. *arXiv preprint arXiv:1804.03599*, 2018.

Salvatore Carta, Anselmo Ferreira, Alessandro Sebastian Podda, Diego Reforgiato Recupero, and Antonio Sanna. Multi-dqn: An ensemble of deep q-learning agents for stock market forecasting. *Expert systems with applications*, 164:113820, 2021.

D. Cer, Y. Yang, S. Kong, N. Hua, N. Limtiaco, R. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, et al. Universal sentence encoder. *arXiv preprint arXiv:1803.11175*, 2018.

Y. Chen, V. Li, K. Cho, and S. Bowman. A stable and effective learning strategy for trainable greedy decoding. *arXiv preprint arXiv:1804.07915*, 2018.

J. Chien and C. Kuo. Markov recurrent neural network language model. In *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 807–813. IEEE, 2019.

C. Danescu-Niculescu-Mizil and L. Lee. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. *arXiv preprint arXiv:1106.3077*, 2011.

S. Dathathri, A. Madotto, J. Lan, J. Hung, E. Frank, P. Molino, J. Yosinski, and R. Liu. Plug and play language models: A simple approach to controlled text generation. *arXiv preprint arXiv:1912.02164*, 2019.

A. Van den Oord, Y. Li, and O. Vinyals. Representation learning with contrastive predictive coding. *arXiv e-prints*, pages arXiv–1807, 2018.

J. Devlin, M. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

M. Fatemi, L. Asri, H. Schulz, J. He, and K. Suleman. Policy networks with two-stage training for dialogue systems. *arXiv preprint arXiv:1606.03152*, 2016.

J. Ficler and Y. Goldberg. Controlling linguistic style aspects in neural language generation. *arXiv preprint arXiv:1707.02633*, 2017.

L. Floridi and M. Chiriatti. GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4):681–694, 2020.

M. Gašić, F. Jurčíček, B. Thomson, K. Yu, and S. Young. On-line policy optimisation of spoken dialogue systems via live interaction with human subjects. In *2011 IEEE Workshop on Automatic Speech Recognition & Understanding*, pages 312–317. IEEE, 2011.

A. Ghandeharioun, J. Shen, N. Jaques, C. Ferguson, N. Jones, A. Lapedriza, and R. Picard. Approximating interactive human evaluation with self-play for open-domain dialog systems. *Advances in Neural Information Processing Systems*, 32, 2019.

E. Greensmith, P. Bartlett, and J. Baxter. Variance reduction techniques for gradient estimates in reinforcement learning. *Journal of Machine Learning Research*, 5(9), 2004.

J. Gu, K. Cho, and V. Li. Trainable greedy decoding for neural machine translation. *arXiv preprint arXiv:1702.02429*, 2017.

T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *ICML*, 2018.

D. Hafner, T. Lillicrap, M. Norouzi, and J. Ba. Mastering atari with discrete world models. *arXiv preprint arXiv:2010.02193*, 2020.

B. Hancock, A. Bordes, P. Mazare, and J. Weston. Learning from dialogue after deployment: Feed yourself, chatbot! *arXiv preprint arXiv:1901.05415*, 2019.

A. Holtzman, J. Buys, M. Forbes, A. Bosselut, D. Golub, and Y. Choi. Learning to write with cooperative discriminators. *arXiv preprint arXiv:1805.06087*, 2018.

N. Hurley and S. Rickard. Comparing measures of sparsity. *IEEE Transactions on Information Theory*, 55(10):4723–4741, 2009.

Youngsoo Jang, Jongmin Lee, and Kee-Eung Kim. Gpt-critic: Offline reinforcement learning for end-to-end task-oriented dialogue systems. In *International Conference on Learning Representations*, 2021.

N. Jaques, A. Ghandeharioun, J. Shen, C. Ferguson, A. Lapedriza, N. Jones, S. Gu, and R. Picard. Way off-policy batch deep reinforcement learning of implicit human preferences in dialog. *arXiv:1907.00456*, 2019.

E. Kharitonov, M. Baroni, and D. Hupkes. How bpe affects memorization in transformers. *arXiv preprint arXiv:2110.02782*, 2021.

Roger Koenker and Kevin F Hallock. Quantile regression. *Journal of economic perspectives*, 15(4): 143–156, 2001.

Ilya Kostrikov, Ashvin Nair, and Sergey Levine. Offline reinforcement learning with implicit q-learning. *arXiv preprint arXiv:2110.06169*, 2021.

T. Kulkarni, K. Narasimhan, A. Saeedi, and J. Tenenbaum. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. *Advances in neural information processing systems*, 29, 2016.

A. Kumar, A. Zhou, G. Tucker, and S. Levine. Conservative q-learning for offline reinforcement learning. *Advances in Neural Information Processing Systems*, 33:1179–1191, 2020.

E. Levin and R. Pieraccini. A stochastic model of computer-human interaction for learning dialogue strategies. In *Eurospeech*, volume 97, pages 1883–1886. Citeseer, 1997.

J. Li, M. Galley, C. Brockett, J. Gao, and B. Dolan. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*, 2015.

J. Li, W. Monroe, A. Ritter, M. Galley, J. Gao, and D. Jurafsky. Deep reinforcement learning for dialogue generation. *arXiv preprint arXiv:1606.01541*, 2016.

J. Li, W. Monroe, T. Shi, S. Jean, A. Ritter, and D. Jurafsky. Adversarial learning for neural dialogue generation. *arXiv preprint arXiv:1701.06547*, 2017.

Z. Li, J. Kiseleva, and M. De Rijke. Dialogue generation: From imitation learning to inverse reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6722–6729, 2019.

W. Liao, B. Zeng, X. Yin, and P. Wei. An improved aspect-category sentiment analysis model for text sentiment analysis based on roberta. *Applied Intelligence*, 51(6):3522–3533, 2021.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

Bo Pang, Erik Nijkamp, Wenjuan Han, Linqi Zhou, Yixian Liu, Kewei Tu, et al. Towards holistic and automatic evaluation of open-domain dialogue generation. 2020.

Y. Park, J. Cho, and G. Kim. A hierarchical latent structure for variational conversation modeling. *arXiv preprint arXiv:1804.03424*, 2018.

B. Peng, C. Zhu, C. Li, X. Li, J. Li, M. Zeng, and J. Gao. Few-shot natural language generation for task-oriented dialog. *arXiv preprint arXiv:2002.12328*, 2020.

A. Razavi, A. van den Oord, and O. Vinyals. Generating diverse high-resolution images with vq-vae. 2019.

A. Saleh, N. Jaques, A. Ghandeharioun, J. Shen, and R. Picard. Hierarchical reinforcement learning for open-domain dialog. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8741–8748, 2020.

C. Sankar, S. Subramanian, C. Pal, S. Chandar, and Y. Bengio. Do neural dialog systems use the conversation history effectively? an empirical study. *arXiv preprint arXiv:1906.01603*, 2019.

I. Serban, A. Sordoni, Y. Bengio, A. Courville, and J. Pineau. Building end-to-end dialogue systems using generative hierarchical neural network models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

I. Serban, C. Sankar, M. Germain, S. Zhang, Z. Lin, S. Subramanian, T. Kim, M. Pieper, S. Chandar, N. Ke, et al. A deep reinforcement learning chatbot. *arXiv preprint arXiv:1709.02349*, 2017a.

I. Serban, A. Sordoni, R. Lowe, L. Charlin, J. Pineau, A. Courville, and Y. Bengio. A hierarchical latent variable encoder-decoder model for generating dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017b.

P. Shah, D. Hakkani-Tur, B. Liu, and G. Tür. Bootstrapping a neural conversational agent with dialogue self-play, crowdsourcing and on-line reinforcement learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers)*, pages 41–51, 2018.

W. Shi and Z. Yu. Sentiment adaptive end-to-end dialog systems. *arXiv preprint arXiv:1804.10731*, 2018.

J. Shin, P. Xu, A. Madotto, and P. Fung. Generating empathetic responses by looking ahead the user's sentiment. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7989–7993. IEEE, 2020.

R. Shu, T. Nguyen, Y. Chow, T. Pham, K. Than, M. Ghavamzadeh, S. Ermon, and H. Bui. Predictive coding for locally-linear control. In *International Conference on Machine Learning*, pages 8862–8871. PMLR, 2020.

S. Singh, D. Litman, M. Kearns, and M. Walker. Optimizing dialogue management with reinforcement learning: Experiments with the njfun system. *Journal of Artificial Intelligence Research*, 16:105–133, 2002.

N. Subramani, S. Bowman, and K. Cho. Can unconditional language models recover arbitrary sentences? *Advances in Neural Information Processing Systems*, 32, 2019.

I. Sutskever, O. Vinyals, and Q. Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014.

R. Sutton, D. McAllester, S. Singh, and Y. Mansour. Policy gradient methods for reinforcement learning with function approximation. *Advances in neural information processing systems*, 12, 1999.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

S. Verma, J. Fu, M. Yang, and S. Levine. Chai: A chatbot ai for task-oriented dialogue with offline reinforcement learning. *arXiv preprint arXiv:2204.08426*, 2022.

M. Walker. An application of reinforcement learning to dialogue strategy selection in a spoken dialogue system for email. *Journal of Artificial Intelligence Research*, 12:387–416, 2000.

W. Wei, Q. Le, A. Dai, and J. Li. Airdialogue: An environment for goal-oriented dialogue research. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3844–3854, 2018.

J. Williams and S. Young. Partially observable markov decision processes for spoken dialog systems. *Computer Speech & Language*, 21(2):393–422, 2007.

M. Yang and O. Nachum. Representation matters: Offline pretraining for sequential decision making. In *International Conference on Machine Learning*, pages 11784–11794. PMLR, 2021.

Denis Yarats and Mike Lewis. Hierarchical text generation and planning for strategic dialogue. In *International Conference on Machine Learning*, pages 5591–5599. PMLR, 2018.

S. Young, M. Gašić, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu. The hidden information state model: A practical framework for pomdp-based spoken dialogue management. *Computer Speech & Language*, 24(2):150–174, 2010.

Y. Zhang, S. Sun, M. Galley, Y. Chen, C. Brockett, X. Gao, J. Gao, J. Liu, and B. Dolan. Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv preprint arXiv:1911.00536*, 2019.

T. Zhao, K. Xie, and M. Eskenazi. Rethinking action spaces for reinforcement learning in end-to-end dialog agents with latent variable models. *arXiv preprint arXiv:1902.08858*, 2019.

L. Zhou, J. Gao, D. Li, and H. Shum. The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, 46(1):53–93, 2020.

D. Ziegler, N. Stiennon, J. Wu, T. Brown, A. Radford, D. Amodei, P. Christiano, and G. Irving. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*, 2019.

A ADDITIONAL RESULTS

A.1 DIALOGUE MANAGEMENT RESULTS ON CORNELL MOVIE CORPUS

Method	Diversity	Dist-1	Dist-2	Dist-3	Perplexity
MoE-1	0.071 ± 0.05	0.23	0.55	0.62	69.14 ± 37.43
MoE-2	0.12 ± 0.04	0.31	0.60	0.79	**43.87 ± 28.81**
MoE-3	0.11 ± 0.06	0.25	0.67	0.70	65.55 ± 46.64
MoE-4	**0.17 ± 0.04**	**0.36**	**0.73**	**0.81**	48.12 ± 39.21
Trans.	0.060 ± 0.05	0.27	0.50	0.66	71.33 ± 40.14
VHRED	0.091 ± 0.03	0.26	0.49	0.61	122.48 ± 55.13

Table 6: Accuracy (Perplexity) and Diversity of Language Primitive Experts Trained with Cornell.

Method	User Tot. Sent.	User Sent. Trans.	Perplexity
MoE-2 IQL	**1.455 ± 0.23**	**0.80 ± 0.28**	44.66 ± 28.28
MoE-2 Ens-Q	0.34 ± 0.25	0.18 ± 0.26	**42.05 ± 26.79**
MoE-2 KLC	0.267 ± 0.36	0.15± 0.28	48.88 ± 10.47
VHRL	0.343 ± 0.27	0.13 ± 0.19	56.51 ± 29.64
VHRL-KLC	0.521 ± 0.31	0.52 ± 0.24	68.53 ± 37.92
VHRL-SAC	0.28 ± 0.33	0.15± 0.28	60.11 ± 40.67

Table 7: Quality of MBRL-based DM Trained with Cornell Dataset.

Method	Question	Exploration	Positive Sent.	Negative Sent.	Sent. Coherence	Joy	Optimism	Anger	Sadness
MoE-1	1.27 ± 0.33	0.30 ± 0.20	1.28 ± 0.51	1.21 ± 0.80	0.75 ± 0.49	1.99 ± 0.51	-0.05 ± 0.94	2.03 ± 0.67	3.20 ± 0.68
MoE-2	**1.58 ± 0.39**	**0.33 ± 0.17**	**3.55 ± 0.99**	1.90 ± 0.5	0.69 ± 0.40	**2.44 ± 0.71**	**2.11 ± 0.99**	**2.71 ± 0.69**	3.45 ± 0.83
MoE-3	1.05 ± 0.50	0.21 ± 0.25	0.30 ± 0.39	1.50 ± 0.61	0.71 ± 0.29	0.73 ± 0.62	0.20 ± 0.60	0.48 ± 0.25	2.17 ± 0.74
MoE-4	1.51 ± 0.28	0.27 ± 0.41	1.02 ± 0.99	**2.25 ± 0.80**	0.76 ± 0.41	0.99 ± 0.74	0.62 ± 0.47	2.69 ± 0.65	**3.50 ± 0.91**
WD	0.84 ± 0.83	0.20 ± 0.19	-0.51 ± 0.25	0.57 ± 0.88	**0.80 ± 0.18**	-0.12 ± 0.75	-1.63 ± 0.42	0.78 ± 0.53	1.19 ± 0.51
PPLM	0.44 ± 0.26	0.29 ± 0.20	-0.26 ± 0.83	0.42 ± 1.03	0.65 ± 0.32	-0.75 ± 0.71	-0.52 ± 0.49	0.61 ± 0.27	1.02 ± 0.77
Trans. RL*	2.20 ± 0.32	0.44 ± 0.26	3.58 ± 0.73	2.98 ± 0.79	0.82 ± 0.30	3.06 ± 0.91	1.64 ± 0.41	3.22 ± 0.88	3.98 ± 0.84

Table 8: Quality of Each Expert PPLM Trained on Cornell Dataset w.r.t. Its Trained Label.

A.2 OFFLINE RL DIALOGUE MANAGEMENT RESULTS ON REDDIT CASUAL AND CORNELL

Method	User Tot. Sent.	User Sent. Trans.	Perplexity
MoE-4 IQL	**0.904 ± 0.58**	**0.68 ± 0.37**	39.89 ± 27.34
MoE-4 Ens-Q	0.183 ± 0.5	0.14 ± 0.30	41.10 ± 29.91
MoE-4 KLC	0.273 ± 0.37	0.19 ± 0.29	**32.21 ± 18.65**
VHRL	0.4055 ± 0.43	0.25 ± 0.32	35.89 ± 18.51
VHRL-KLC	0.198 ± 0.37	0.14 ± 0.19	43.45 ± 29.13
VHRL-SAC	0.132 ± 0.43	0.07 ± 0.14	41.17 ± 26.42

Table 9: Quality of Offline RL-based DM Trained with Reddit Casual Dataset.

Method	User Tot. Sent.	User Sent. Trans.	Perplexity
MoE-2 IQL	-1.298 ± 0.24	-0.10 ± 0.25	40.17 ± 29.47
MoE-2 Ens-Q	-1.557 ± 0.27	-0.18 ± 0.22	36.21 ± 20.82
MoE-2 KLC	**-1.255 ± 0.23**	-0.12 ± 0.24	**33.17 ± 25.14**
VHRL	-1.319 ± 0.32	-0.11 ± 0.20	44.82 ± 32.11
VHRL-KLC	-1.545 ± 0.28	**-0.04 ± 0.16**	57.26 ± 35.361
VHRL-SAC	-1.482 ± 0.33	-0.09 ± 0.15	65.41 ± 43.25

Table 10: Quality of Offline RL-based DM Trained with Cornell Dataset.

A.3 Sample Utterances for EXP 1

A.3.1 Reddit Casual

1. **Input Conversation:** (...) i mean , i have n't changed or done anything drastically different , but things have been good .
    - **MoE-1 R1:** but my outlook on life is n't meh .
    - **MoE-1 R2:** but my outlook on life is subjective .
    - **MoE-1 R3:** but my outlook on life is still meh .
    - **MoE-1 R4:** but my view on life is still meh ?
    - **MoE-1 R5:** but my outlook on life is n't meh ?

    - **MoE-2 R1:** i love the good things that i 'm starting too.
    - **MoE-2 R2:** i do love your life , because mine 's nothing .
    - **MoE-2 R3:** no , i 'm not too busy in life
    - **MoE-2 R4:** good for you , i 'm working on it .
    - **MoE-2 R5:** but my life is boring and boring .

    - **MoE-3 R1:** but my outlook on life is still sad .
    - **MoE-3 R2:** but my outlook on life is the same .
    - **MoE-3 R3:** but my outlook on life is my same !
    - **MoE-3 R4:** but my outlook on life is not good .
    - **MoE-3 R5:** but my outlook on life is my friend .

    - **MoE-4 R1:** i 'm in my early 20 's , i do n't know what to do
    - **MoE-4 R2:** i 'm so happy for you .
    - **MoE-4 R3:** for me , it 's tough to go through life with pain .
    - **MoE-4 R4:** i used to think that you 'd enjoy that much more than most people .
    - **MoE-4 R5:** can you describe what happened ?

2. **Input Conversation:** (...) but my outlook on life is still meh
    - **MoE-1 R1:** no need for drastically change, just take it slow for a break ?
    - **MoE-1 R2:** no need to slow down, just take it a long day ?
    - **MoE-1 R3:** no need to slow down , just take a step too long .
    - **MoE-1 R4:** no need a blanket , just take it off .
    - **MoE-1 R5:** no need for a breath , just take it a bit of time

    - **MoE-2 R1:** being a good person is n't a bad thing
    - **MoE-2 R2:** i 'm in a similar situation for it .
    - **MoE-2 R3:** take a break , try to find something else .
    - **MoE-2 R4:** oh yeah , life 's pretty good for me .
    - **MoE-2 R5:** like you 're in a bad place for a few years ?

    - **MoE-3 R1:** no need for drastic changes , just make it super busy
    - **MoE-3 R2:** no need for drastic changes , just do it extra time
    - **MoE-3 R3:** no need for unnecessary stress , just move the next direction
    - **MoE-3 R4:** no change for unnecessary things , just move on .
    - **MoE-3 R5:** no need for unnecessary consumption , just take with the move

    - **MoE-4 R1:** it seems like a great time to just go outside and play a few songs , i 'm always glad to hear that you have a good friend
    - **MoE-4 R2:** i know life 's not perfect , it can be difficult at times !
    - **MoE-4 R3:** end complaining all the time and just consider your life a failure .
    - **MoE-4 R4:** you 'd be surprised at how nice life can go !
    - **MoE-4 R5:** i mean some sort of hobby can definitely help out .

3. **Input Conversation:** (...) no need for drastically difference , just take it slow .
    - **MoE-1 R1:** and it 's nice to hear someone else 's good , but it 's all about my life .
    - **MoE-1 R2:** and it 's nice to hear , when you have new seasons of american world is just a nature in canada .
    - **MoE-1 R3:** and it 's nice to hear why you 've been doing it , but my friend is still in it ...
    - **MoE-1 R4:** and it 's very good to hear why you 've been good with my wife , a lot of things that 's happening in the same way .
    - **MoE-1 R5:** and it 's nice to hear things have been good , my view of life changes with time and mood

    - **MoE-2 R1:** i find it hard to get out of this world , but there is also a good way to find a positive way of life .,
    - **MoE-2 R2:** you seem to be in the same situation , but it is a good way to do it for a lot of things in life
    - **MoE-2 R3:** my life is n't bad , i 'm in a very long time and i m pretty happy about it .
    - **MoE-2 R4:** i just wanted to see what 's happening in my life and i m not sure what to do .
    - **MoE-2 R5:** at this point i 'm pretty sure the most part of my life is pretty normal , but it s a lot of positive .

    - **MoE-3 R1:** and it 's nice to hear that things have been good for you that your view life is a lot like mine .
    - **MoE-3 R2:** and it 's nice to hear that things have been good for you so my perspective of life is too much and so i think that 's a pass
    - **MoE-3 R3:** and it 's nice to hear that things have been good for you in my view of life , like a lot of people are trying to have
    - **MoE-3 R4:** and it 's nice to hear that things have been good for you , my view my life is awful that , but i feel like this is a lot better
    - **MoE-3 R5:** and it 's nice to hear that things have been good for you in the future of your life , so maybe you 're looking forward to it

    - **MoE-4 R1:** i 've been trying to learn things about myself .
    - **MoE-4 R2:** that 's like i would be at this stage of life .
    - **MoE-4 R3:** you will need to enjoy life for yourself .
    - **MoE-4 R4:** for me it 's like i do n't have the time .
    - **MoE-4 R5:** but like you said , you 've been trying to get your life together and learn to be a normal person to love yourself .

4. **Input Conversation:** (...) and it 's nice to hear that things have been good for you d my view of life is meh too , it changes depending on time and mood
    - **MoE-1 R1:** and i do n't know why i say this , this is not a relevant post , but i can really say the same things .
    - **MoE-1 R2:** and i do n't know why i say this , this is not even relevant , but any things will be good .
    - **MoE-1 R3:** and i do n't know why i say this , this is not even a relevant topic , i do n't remember the same things i said
    - **MoE-1 R4:** and i do n't know why i say this , this is n't even relevant , but definitely i enjoy the same things
    - **MoE-1 R5:** and i do n't know why i say this , this is not a relevant topic , but i think there 's actually a bunch of things to do

    - **MoE-2 R1:** i think i 'm just trying to be a good person , but i m not sure what it is like in my life .
    - **MoE-2 R2:** for me it 's my life i feel like it s the same way of being around the world and i m pretty sure it s not a lot

- **MoE-2 R3:** to be honest , i 'm not sure what the most beautiful things are like in the world
- **MoE-2 R4:** i feel like you 're going to be a good person , but i 'm not a very happy about the world of things you are around
- **MoE-2 R5:** i just want to be in a way to see my life and i m not really sure how much of other things i like

- **MoE-3 R1:** and i do n't know why i say this , this is not really relevant , but i do wish you the best .
- **MoE-3 R2:** and i do n't know why i say this , this is not really relevant , but i can count the experience
- **MoE-3 R3:** and i do n't know why i say this , this is not even relevant , but now i feel like it 's the kind of boring ...
- **MoE-3 R4:** and i do n't know why i say this , this is not even relevant , but i like watching ultra match in my room .
- **MoE-3 R5:** and i do n't know why i say this , this is not even relevant , but i do feel like you 're welcome here .

- **MoE-4 R1:** i hope things get better for you soon
- **MoE-4 R2:** i think it 's fine , i enjoy life and could be a little happier every day
- **MoE-4 R3:** i like it d be a good time to think about the things that you don t need to do .
- **MoE-4 R4:** but let 's forget on those things that make you miserable and trying to focus on the basics .
- **MoE-4 R5:** i know it 's not perfect , but enjoy the things you have .

5. **Input Conversation:** (...) and i do n't know why i say this , this is not even relevant , but remember rule 32 enjoy the little things
    - **MoE-1 R1:** i 'm trying to enjoy the little things like walking down the street , or playing random games .
    - **MoE-1 R2:** i 'm just trying to enjoy the little things like sports , like small talk , or random ones , or even give up
    - **MoE-1 R3:** i 'm just trying to enjoy the little things like walking down the street with the same interactions or drinking
    - **MoE-1 R4:** i 'm trying to enjoy the little things like going down the beach with the same thing , or a nice spot
    - **MoE-1 R5:** i 'm trying to enjoy the little things like walking down the beach , watching the weird movies , listen to random music

    - **MoE-2 R1:** this is a good time , i 'm just looking for a place to go and i 've always wanted to go to the store
    - **MoE-2 R2:** my favorite thing is that i 'm just looking for a good thing to do , but i 'm always looking forward to the people i 've been
    - **MoE-2 R3:** i might be in the same boat , but i 'm a little scared to know what is a really good thing to look at the moment
    - **MoE-2 R4:** i 've been trying to find out what i want to do , but i 'm not really sure
    - **MoE-2 R5:** i know it 's a good thing , but i 'm always in the same boat as i 've been in a long run

    - **MoE-3 R1:** yeah haha , i 'm definitely trying to enjoy the little things like walking down the street , or just have a hot cup of tea
    - **MoE-3 R2:** yeah haha , i 'm definitely trying to enjoy the little things like walking down the street or seeing what is in my neighborhood
    - **MoE-3 R3:** yes , i m definitely trying to enjoy the little things like walking down stairs , but i can t figure out what to see
    - **MoE-3 R4:** yeah dude , i m definitely trying to enjoy the little things like walking down the street with a huge fan of these kids

- **MoE-3 R5:** yeah man , i ’m definitely trying to enjoy the little things like from walking up the street or to my little kid ’s house

- **MoE-4 R1:** that ’s what i ’m really trying to do and enjoy the moment .
- **MoE-4 R2:** i ’ve always been like that , i ’ve been really chill at least trying to enjoy my life
- **MoE-4 R3:** i ’m trying to get a good life and have a good time , i ’m not sure what to do in my free time
- **MoE-4 R4:** and i ’ve definitely learned a lot of things to do
- **MoE-4 R5:** i d say that it has been given me both the time and pleasure, but i m still an introvert with few hobbies

A.3.2 Cornell Movie Corpus

1. **Input Conversation:** (...) excuse me
   - **MoE-1 R1:** i do n’t know .
   - **MoE-1 R2:** i do n’t know .
   - **MoE-1 R3:** yes .
   - **MoE-1 R4:** what do you mean ?
   - **MoE-1 R5:** i ’m sorry .

   - **MoE-2 R1:** i ’ll give you a few minutes .
   - **MoE-2 R2:** i do n’t want to talk about it .
   - **MoE-2 R3:** may i come in ?
   - **MoE-2 R4:** where is this , mr. kessler ?
   - **MoE-2 R5:** what do you think ?

   - **MoE-3 R1:** what ’s it ? are you alright ?
   - **MoE-3 R2:** you want me to read ?
   - **MoE-3 R3:** i do n’t know .
   - **MoE-3 R4:** i ’ll just have to kill you .
   - **MoE-3 R5:** you ca n’t say it ’s over .

   - **MoE-4 R1:** well if you hang up , you ’ll die on it .
   - **MoE-4 R2:** superman ? oh , was he here ?
   - **MoE-4 R3:** well , you know , that ’s a lie , i do n’t have your money .
   - **MoE-4 R4:** are you the killer ben ?
   - **MoE-4 R5:** there ’s nothing to be wrong with that .

2. **Input Conversation:** (...) yes ?
   - **MoE-1 R1:** what do you mean ?
   - **MoE-1 R2:** i ’m not sure .
   - **MoE-1 R3:** what ’s the matter ?
   - **MoE-1 R4:** i do n’t know .
   - **MoE-1 R5:** no .

   - **MoE-2 R1:** what did you say ?
   - **MoE-2 R2:** he says that you ’re not a good man .
   - **MoE-2 R3:** i ’ll tell you what i ’ll do . i ’ll be right back .
   - **MoE-2 R4:** do you want to go to the police ?
   - **MoE-2 R5:** are you still in love with him ?

   - **MoE-3 R1:** sam , i ’m getting married .
   - **MoE-3 R2:** johnny . it was announced this morning . he ’s going to play the lead in the new film .

- **MoE-3 R3:** i 'm sorry , mister mcmurphy , i count only eight .
- **MoE-3 R4:** i said i was n't looking for him . will you excuse my miss ... ?
- **MoE-3 R5:** pay ! in and i say !

- **MoE-4 R1:** i would like to marry you . but i have already promised myself ... to god .
- **MoE-4 R2:** do you know where he 's living ?
- **MoE-4 R3:** i know you 're soft .
- **MoE-4 R4:** it 's susan 's cooking , claire . why uncle james lets them stay ?
- **MoE-4 R5:** i do n't want to be any trouble .

3. **Input Conversation:** (...) can i ask my friend , can he come in ? just for a few minutes ?
    - **MoE-1 R1:** i 'm not sure .
    - **MoE-1 R2:** i do n't know
    - **MoE-1 R3:** he 's not a man , is n't he ?
    - **MoE-1 R4:** no, i 'm sorry .
    - **MoE-1 R5:** what do you mean ?

    - **MoE-2 R1:** all right . what are you going to do ?
    - **MoE-2 R2:** no , i 've never seen him before .
    - **MoE-2 R3:** i want to be with him .
    - **MoE-2 R4:** this is a very interesting man .
    - **MoE-2 R5:** he 's not the greatest . you do n't know .

    - **MoE-3 R1:** oh , come on , i 'm sick and tired of your life .
    - **MoE-3 R2:** no . i still do n't have any reason to do that .
    - **MoE-3 R3:** my god , are you kidding , i 'm a little tired .
    - **MoE-3 R4:** i do n't understand . where is he ?
    - **MoE-3 R5:** no . i still do n't have any reason let you be together

    - **MoE-4 R1:** oh , of course !
    - **MoE-4 R2:** because he was scary , right ?
    - **MoE-4 R3:** i 'm sure you 're a good friend .
    - **MoE-4 R4:** can i get you a glass of champagne ?
    - **MoE-4 R5:** so i like to make a point . you got a fucking friend that does n't work

4. **Input Conversation:** (...) your friend ?
    - **MoE-1 R1:** i do n't know
    - **MoE-1 R2:** you 're not going to tell him ?
    - **MoE-1 R3:** well , i 'm not sure .
    - **MoE-1 R4:** i do n't know .
    - **MoE-1 R5:** i 'm sorry

    - **MoE-2 R1:** i like him . he 's a very good man .
    - **MoE-2 R2:** maybe i should go home and talk to him .
    - **MoE-2 R3:** i 'm not going to let him go .
    - **MoE-2 R4:** oh , please .
    - **MoE-2 R5:** if you do n't mind , you 'll have to do it , will you ?

    - **MoE-3 R1:** what 's the idea of working with you ? what 's the matter ? !
    - **MoE-3 R2:** i 'll look into her . tell 'em .
    - **MoE-3 R3:** that 's something i have n't done the before .
    - **MoE-3 R4:** so , our lawyers are here .
    - **MoE-3 R5:** we 've got some plans here .

- **MoE-4 R1:** my friend , you must be really glad to help . if you do n't mind .
- **MoE-4 R2:** is n't there something you can do . how can you help at night ?
- **MoE-4 R3:** how do you do , mrs. ryan ? i 'd heard you were beautiful , and so you are .
- **MoE-4 R4:** the winning side would have paid you much better .
- **MoE-4 R5:** now , here 's a babe ! I give you a blue ox , you got ta help me !

5. **Input Conversation:** (...) he 's going back to the front this evening . i ca n't see him otherwise .

   - **MoE-1 R1:** i 'm not going to be here .
   - **MoE-1 R2:** what do you mean ?
   - **MoE-1 R3:** i 'm not sure . it 's not your fault .
   - **MoE-1 R4:** i do n't know .
   - **MoE-1 R5:** you 're not going to tell him ?

   - **MoE-2 R1:** i think i 'll tell you what i 'm gon na do .
   - **MoE-2 R2:** he 's on the way , right ?
   - **MoE-2 R3:** that 's not true .
   - **MoE-2 R4:** not too late . i 'm not going to be here .
   - **MoE-2 R5:** what are you talking about ? he 's a very sweet man .

   - **MoE-3 R1:** what ? i think it 's wonderful .
   - **MoE-3 R2:** sure , i guess !
   - **MoE-3 R3:** oh , for christ 's sake .
   - **MoE-3 R4:** they 'll be all right , you 're my guest .
   - **MoE-3 R5:** leon , you are not going to negotiate !

   - **MoE-4 R1:** ok. he asks about you all the time . i want to tell you .
   - **MoE-4 R2:** the way i feel , believe me , it 's hard to part and see a guy 's picture .
   - **MoE-4 R3:** do n't worry . i 'll take care of it .
   - **MoE-4 R4:** certainly it 's important .
   - **MoE-4 R5:** yeah , and what a real hero .

A.4 SAMPLE UTTERANCES FOR EXP 2: QUESTION INTENT

A.4.1 REDDIT CASUAL

1. **Sample Conversation:** (...) i love how genuine your comments are and you only comment if it s how you really feel and not for karma like when you wish people a merry christmas

   - **MoE-1 R1:** what was your favorite guys and why did you have a boyfriend ?
   - **MoE-1 R2:** i really love your username , and i am so happy for you .
   - **MoE-1 R3:** what was your favorite guys and why did you have a nice day ?

   - **MoE-2 R1:** how did you find someone that you like ?
   - **MoE-2 R2:** just wondering if you re interested in reddit ?
   - **MoE-2 R3:** how did you feel about it ?

   - **MoE-3 R1:** people like you really have a special place in my heart .
   - **MoE-3 R2:** people like you really have a special place in my heart to me
   - **MoE-3 R3:** people like you really have a special place in my heart?

   - **MoE-2 R1:** how are you liking it ?
   - **MoE-2 R2:** do you know i just love being together with you ?
   - **MoE-2 R3:** how are you my friend ? ?

2. **Sample Conversation:** (...) people like you really have a special place in my heart .
    - **MoE-1 R1:** that s nice !
    - **MoE-1 R2:** happy cake day !
    - **MoE-1 R3:** that sounds nice , you have a good time to stay alone ?

    - **MoE-2 R1:** just wondering if you re happy with me ?
    - **MoE-2 R2:** how did you feel about it ?
    - **MoE-2 R3:** how did you get this ?

    - **MoE-3 R1:** hope your christmas was a nice one to celebrate with you
    - **MoE-3 R2:** hope your christmas was pretty good , my friend being celebrate with .
    - **MoE-3 R3:** hope your christmas was nice with my friend who celebrate with you

    - **MoE-4 R1:** omg why !
    - **MoE-4 R2:** how far into your heart ?
    - **MoE-4 R3:** how are you liking this happiness as well ?

3. **Sample Conversation:** (...) hope your christmas was pretty good , my friend assuming you celebrate .
    - **MoE-1 R1:** thank you !
    - **MoE-1 R2:** thanks !,
    - **MoE-1 R3:** thank you !

    - **MoE-2 R1:** how did you guys meet this year ?
    - **MoE-2 R2:** how was your day going ?
    - **MoE-2 R3:** how did you celebrate this year ?

    - **MoE-3 R1:** haha thank you !
    - **MoE-3 R2:** thank you !
    - **MoE-3 R3:** thank you .

    - **MoE-4 R1:** how 's your day going ?
    - **MoE-4 R2:** how are you liking it so far ?
    - **MoE-4 R3:** i have no idea what it 's about ?

A.4.2 CORNELL MOVIE CORPUS

1. **Sample Conversation:** that 's some face they left hanging on you . you look like ...
    - **MoE-1 R1:** what ? what are you talking about ?
    - **MoE-1 R2:** what ? what are you talking about ?
    - **MoE-1 R3:** what ?

    - **MoE-2 R1:** what ? what 's wrong with it ?
    - **MoE-2 R2:** what ? you think they 're going to kill me ?
    - **MoE-2 R3:** what is it ?

    - **MoE-3 R1:** well . what 's the difference ?
    - **MoE-3 R2:** what ? what 's the deal ?
    - **MoE-3 R3:** what ? what ? what 's wrong with what happened ?

    - **MoE-4 R1:** what ? there 's no pain in my mind ?
    - **MoE-4 R2:** what ? why do you think i want to know ?
    - **MoE-4 R3:** what can i do ? there 's no way i can change .

2. **Sample Conversation:** what hotel ?
    - **MoE-1 R1:** what the fuck are you doing here ?
    - **MoE-1 R2:** what ?
    - **MoE-1 R3:** what are you talking about ?,

    - **MoE-2 R1:** what 's the matter ?
    - **MoE-2 R2:** what about it ?
    - **MoE-2 R3:** who 're you traveling with?

    - **MoE-3 R1:** whatever . i 'm going to work here .
    - **MoE-3 R2:** tom . how are you ?
    - **MoE-3 R3:** what ? what about the hotel ?

    - **MoE-4 R1:** do you think i want the plaza ?
    - **MoE-4 R2:** what do you think ? where 's my mother ?
    - **MoE-4 R3:** what hotel i should go?

3. **Sample Conversation:** the holiday hotel .
    - **MoE-1 R1:** what are you talking about ?
    - **MoE-1 R2:** what ?
    - **MoE-1 R3:** what are you doing here ?

    - **MoE-2 R1:** what is your travel plan?
    - **MoE-2 R2:** how can i get there?
    - **MoE-2 R3:** why do you want to be there?

    - **MoE-3 R1:** what ? are you walking in there ?,
    - **MoE-3 R2:** absolutely . what 's up with the hotel ?
    - **MoE-3 R3:** what ? bring 'em in

    - **MoE-4 R1:** what are you ? a cop ?
    - **MoE-4 R2:** what about that ? i work here .
    - **MoE-4 R3:** what ? ... oh you know he 's there ?,

A.5 Sample Utterances for EXP 2: Positive Sentiment Intent

A.5.1 Reddit Casual

1. **Sample Conversation:** (...) it s definitely not the college experience .
    - **MoE-1 R1:** i feel like i m more of a high school and talking to people .
    - **MoE-1 R2:** i feel like i 'm not a teacher , and i talk to , i talk to people about it .
    - **MoE-1 R3:** i feel like a lot of people are just as nervous as i talk to , and i really really like talking about .

    - **MoE-2 R1:** i m so proud of you for being able to keep up with it !
    - **MoE-2 R2:** i hope you re having a great day !
    - **MoE-2 R3:** i have a very good time on social media !

    - **MoE-3 R1:** seems like middle school to end up way more stable than everyone and don t know who does this
    - **MoE-3 R2:** seems like real good compared to everyone after work and nobody is around me even the age of 35
    - **MoE-3 R3:** feels like a real awkward going to say though and nobody cares about each other with me

- **MoE-4 R1:** if you want to get into college , i ’m pretty good at it !
- **MoE-4 R2:** thank you for taking the time to look at it , i ’m trying to get a few more opportunities for you now for your semester
- **MoE-4 R3:** good for you , i like college and have more people in my life , i know better than high school i ve been

2. **Sample Conversation:** (...) feels more like high school except way easier to mess up and nobody really talks to anyone lmao .
   - **MoE-1 R1:** i m on reddit , and i m going to be pretty far from my first year s college , but i d probably have to go to college
   - **MoE-1 R2:** i m not really good at all , but i m a college student , i d be working on a job , i d be able to go,
   - **MoE-1 R3:** i m pretty sure it s a lot easier to get a job , but i guess it s a lot of work in my field

   - **MoE-2 R1:** i m going to be a senior and i m happy to meet people !
   - **MoE-2 R2:** i hope you re having a great time !
   - **MoE-2 R3:** i get to see my new friends and family !

   - **MoE-3 R1:** i ’m looking forward to being able to say , but not because i am a great life all for everyone to see each other once in a while
   - **MoE-3 R2:** i ’m looking forward to going out , not because i feel so much better knowing i will still have a great day , and i ’m happy to see
   - **MoE-3 R3:** i ’m looking forward to going home , not only because i think i will be a lot happier with my family and looking at the very young people .

   - **MoE-4 R1:** for sure , i ve been in college for the past few years that i ve had plenty of people in my life .
   - **MoE-4 R2:** thank you for your input , i appreciate it !
   - **MoE-4 R3:** you ll be proud of yourself and i hope you are able to enjoy those people on your first year

3. **Sample Conversation:** (...) i m looking forward to finally being out , not only because i think i ll be a lot happier at a full uni , but also bc i hate home
   - **MoE-1 R1:** i finally got the job that i m going to be a very good dad , and it just sucks .
   - **MoE-1 R2:** i finally got a new job that i feel like is not a good thing .
   - **MoE-1 R3:** i finally got the hang of my parents , and i feel very lucky to be very far as possible .

   - **MoE-2 R1:** now i m a college student and i m very proud of myself !
   - **MoE-2 R2:** i ll try to meet new people !
   - **MoE-2 R3:** i feel like i m a little too proud of myself !

   - **MoE-3 R1:** yeah well and the whole thing will really help me feel a bit shitty right now is n’t easy
   - **MoE-3 R2:** yeah well this is the start and really make me feel good a bit better than what i am .
   - **MoE-3 R3:** yeah well and the goal is really help me feel good since everyone is mostly different in their lives .

   - **MoE-4 R1:** good advice , i really like it when i have to get a college degree , but i ’ve also got a decent amount of friends who have not
   - **MoE-4 R2:** thank you for the advice , i really like to keep on going for college and i ’m sure my friends will go to college
   - **MoE-4 R3:** for sure , i want to get back on track to how much i used to be doing , but i don t really like the idea of studying

A.5.2 CORNELL MOVIE CORPUS

1. **Sample Conversation:** (...) no i fucking do n't ! you ca n't make a cup of tea , edward .
    - **MoE-1 R1:** what ?
    - **MoE-1 R2:** no .
    - **MoE-1 R3:** no , i 'm not .

    - **MoE-2 R1:** yeah , i got it .
    - **MoE-2 R2:** oh , i 'm so glad you say it .
    - **MoE-2 R3:** yes , it 's a very nice idea .

    - **MoE-3 R1:** yes , sir . what 's wrong ?
    - **MoE-3 R2:** yes . yes , sir .
    - **MoE-3 R3:** yes . who is this ?

    - **MoE-4 R1:** oh , yes . i 'm going to miss you .,
    - **MoE-4 R2:** yes . i 'm sorry , i did n't know how to do it .
    - **MoE-4 R3:** yes , yes . i 'll tell you something about it .

2. **Sample Conversation:** (...) why not ? the whole of the british empire was built on cups of tea .
    - **MoE-1 R1:** what ?
    - **MoE-1 R2:** no !
    - **MoE-1 R3:** no , no .

    - **MoE-2 R1:** good . i 'm glad you say that .
    - **MoE-2 R2:** yeah , i got it .
    - **MoE-2 R3:** yes , you are right about that .

    - **MoE-3 R1:** yes . yes , sir .
    - **MoE-3 R2:** yes , of course .
    - **MoE-3 R3:** yes . yes . that 's right .

    - **MoE-4 R1:** yes , i 'll tell you what it is .
    - **MoE-4 R2:** yes , yes . i 'll tell you something else !
    - **MoE-4 R3:** yes , i 'm with you .

3. **Sample Conversation:** (...) if you 'll be good enough to remove those newspapers i 'll have a seat .
    - **MoE-1 R1:** no , i 'm not .
    - **MoE-1 R2:** what ?
    - **MoE-1 R3:** no !

    - **MoE-2 R1:** yes , i think it 's a good idea .
    - **MoE-2 R2:** yes , yes . i 'll be in a minute .
    - **MoE-2 R3:** good . yes , you 're not going to have to do it .

    - **MoE-3 R1:** yes . yes . yes . yes .
    - **MoE-3 R2:** what 's the matter with you ?
    - **MoE-3 R3:** yes , ma'am , i do .

    - **MoE-4 R1:** oh , yes ! i 'll do it .
    - **MoE-4 R2:** yes , yes . i 'm sure you must be a very sick man .
    - **MoE-4 R3:** yes , yes . i 'm sure you do ..

A.6 SAMPLE UTTERANCES FOR EXP 2: NEGATIVE SENTIMENT INTENT

A.6.1 REDDIT CASUAL

1. **Sample Conversation:** (...) it s definitely not the college experience .
   - **MoE-1 R1:** i m not much of a sports person , i find that sort of an emotional connection to others .
   - **MoE-1 R2:** i 'm more of a high school student and i m wondering if anyone else to talk to about this .
   - **MoE-1 R3:** i 'm not a social person , but i 'm just really curious how to talk to other people .

   - **MoE-2 R1:** i don t know if i should be going to school right now .
   - **MoE-2 R2:** i don t know how to do it , i m nervous .
   - **MoE-2 R3:** i don t think i m going to be able to do all the things i ve been doing .

   - **MoE-3 R1:** feels like shit not less to lie than excuse myself to just miss class all the time .
   - **MoE-3 R2:** feels like decent becoming middle school system and not close to him anymore anyways because nobody ever mentioned .
   - **MoE-3 R3:** feels like middle school only harder to stay up and nobody really talks about anything except nowadays .

   - **MoE-4 R1:** because i hate making out of my freshman year i was happy about it
   - **MoE-4 R2:** but i hate to admit it , so like this was the first time i even went to college .
   - **MoE-4 R3:** and people are terrible at figuring out what to do , you could n't pay extra for this year !

2. **Sample Conversation:** (...) feels more like high school except way easier to mess up and nobody really talks to anyone lmao .
   - **MoE-1 R1:** i m not really good at all , but i guess i would have to go to a university for a good job .
   - **MoE-1 R2:** so far that i 'm not happy to have my own experience , but i m also going to have a job that i d like to have .
   - **MoE-1 R3:** i m on the same boat as you , but i m also a college student for a few years .

   - **MoE-2 R1:** i m not going to try and start doing anything for the next semester .
   - **MoE-2 R2:** i don t know if it s the same .
   - **MoE-2 R3:** i don t know if it s going to be the same since i m still in college .

   - **MoE-3 R1:** i m looking forward to going back to being alone , but not because i feel like i m a better fit in with the wait til stressful one
   - **MoE-3 R2:** i 'm looking forward to going back home , not only because i feel that there 's a lot of time to do something productive after i get there
   - **MoE-3 R3:** i 'm looking forward to being at school , not only because i think i 'll be a bit happier now and will have to pay for 4 months

   - **MoE-4 R1:** and i hate it , i get bored !
   - **MoE-4 R2:** i just switched because i was a senior student and i know absolutely hate that i had to do
   - **MoE-4 R3:** and i hate talking about it , i 'll get left out

3. **Sample Conversation:** (...) i m looking forward to finally being out , not only because i think i ll be a lot happier at a full uni , but also bc i hate home
   - **MoE-1 R1:** so i finally got a new job that i am not a very good person .
   - **MoE-1 R2:** i 'm not going to be able to find a way to feel better about it , but i just feel better .

- **MoE-1 R3:** i m not the most comfortable person who wants to be so bad about it or something that i should be able to be able to do .

- **MoE-2 R1:** i feel like i m not going to be honest .
- **MoE-2 R2:** i feel like i m not going to be in the same position .
- **MoE-2 R3:** i don t know if it s all about going to school .

- **MoE-3 R1:** yeah well all and the person does really make me feel better for everyone to be so serious about it .
- **MoE-3 R2:** yeah well this and the beat all does seem feel a bit better because everyone is basically just that it is in their ass .
- **MoE-3 R3:** yeah well and the post really not me feel better a lot better since it is not being considered so big .

- **MoE-4 R1:** and i hate waking up , i feel like it all the time .
- **MoE-4 R2:** and i hate to admit it , i do n't know what 's going on .
- **MoE-4 R3:** and i know there are people who can keep up with their homework but i do n't know how to handle them !

A.6.2 CORNELL MOVIE CORPUS

1. **Sample Conversation:** (...) no i fucking do n't ! you ca n't make a cup of tea , edward .
   - **MoE-1 R1:** i 'm sorry .
   - **MoE-1 R2:** i 'm not sure .
   - **MoE-1 R3:** i do n't know , i do n't know .

   - **MoE-2 R1:** what do i have to do ?
   - **MoE-2 R2:** why do n't you just leave it be ?
   - **MoE-2 R3:** you do n't know what you 're talking about .

   - **MoE-3 R1:** i 'm not interested in what people used to do .
   - **MoE-3 R2:** i 'm afraid that 's not up to me
   - **MoE-3 R3:** yes . what 's the problem with it ?

   - **MoE-4 R1:** i do n't want to hear your problems .
   - **MoE-4 R2:** i do n't care . i have a lot of money in my life .
   - **MoE-4 R3:** i do n't want a goddamn cup of tea . i want to go home . i 'll have to be here , okay ?

2. **Sample Conversation:** (...) why not ? the whole of the british empire was built on cups of tea .
   - **MoE-1 R1:** i do n't know .
   - **MoE-1 R2:** i 'm not sure . i 'm not going to be here .
   - **MoE-1 R3:** i do n't know .

   - **MoE-2 R1:** who do you think you are ?
   - **MoE-2 R2:** who cares ? i do n't know what to say .
   - **MoE-2 R3:** do n't be a fool . do n't you know what i 'm talking about .

   - **MoE-3 R1:** yes . what 's more like this ?
   - **MoE-3 R2:** i do n't believe that either .
   - **MoE-3 R3:** that is correct . and how do we decide ?

   - **MoE-4 R1:** i do n't believe in it !
   - **MoE-4 R2:** i do n't want to hear you anymore . you hear me ?

- **MoE-4 R3:** i do n't want to hear it . i know you 're very clever .

3. **Sample Conversation:** (...) if you 'll be good enough to remove those newspapers i 'll have a seat .
    - **MoE-1 R1:** i do n't know .
    - **MoE-1 R2:** i 'm not sure .
    - **MoE-1 R3:** i do n't know .

    - **MoE-2 R1:** i do n't know what you 're talking about .
    - **MoE-2 R2:** who are you ? can you get out of here ?
    - **MoE-2 R3:** do n't be a fool . what do you want me to do ?

    - **MoE-3 R1:** you 're welcome to one of these , whatever .
    - **MoE-3 R2:** what difference does it make ?
    - **MoE-3 R3:** i 'm afraid not . have to ask you a question .

    - **MoE-4 R1:** i do n't want to hear it . i do n't want to hear any more today .
    - **MoE-4 R2:** look at this ! why are you laughing ?
    - **MoE-4 R3:** i do n't care ! i do n't know what 's good with you !

A.7 Sample Utterances for EXP 2: Joy Emotion Intent

A.7.1 Reddit Casual

1. **Sample Conversation:** (...) besides every anime intro and outro song ever at least from the series i like , i 've been getting a kick out of j rock in general .
    - **MoE-1 R1:** i got a song a few months ago and i can relate .
    - **MoE-1 R2:** i got a song from a song that i can recently finish .
    - **MoE-1 R3:** i got a song from a song that has been a really rough time lately .

    - **MoE-2 R1:** what s your favourite anime to listen to ?
    - **MoE-2 R2:** i just enjoy the new anime and i 'm a big fan of it .
    - **MoE-2 R3:** i 'm always a fan of the anime and i 've been listening to it all the lately

    - **MoE-3 R1:** i got very into french music this year which i can relate every year
    - **MoE-3 R2:** i got into love pop music this year , so i can relate with very well .
    - **MoE-3 R3:** i got into love rap recently this year so i can relate with you haha

    - **MoE-4 R1:** haha i 'm into anime !
    - **MoE-4 R2:** haha , anime manga is considered my first language .
    - **MoE-4 R3:** sounds like my favourite anime and manga japan , and they probably are my main thing .

2. **Sample Conversation:** (...) i got deep into idol music this year so i can relate a bit .
    - **MoE-1 R1:** i 've been thinking about the same as a lot of new bands i listen to .
    - **MoE-1 R2:** i just started a new anime where i feel like i 'm into the same genre .
    - **MoE-1 R3:** i just like the new design and see it on youtube .

    - **MoE-2 R1:** <unk> is my favorite song
    - **MoE-2 R2:** i love the song !
    - **MoE-2 R3:** what do you like about the music ?

    - **MoE-3 R1:** i have really watched deep dark <unk> recently .
    - **MoE-3 R2:** yeah i have heard this song called so long

- **MoE-3 R3:** i know just your song i played video games sometimes .

- **MoE-4 R1:** i 'm also loving the music so far .
- **MoE-4 R2:** haha , i love anime manga from the 90 's .
- **MoE-4 R3:** haha yes but for me it 's all anime .

3. **Sample Conversation:** (...) some people have the same username across multiple websites such as steam or xbox .
    - **MoE-1 R1:** people might even get mad at them , because they are on their own .
    - **MoE-1 R2:** i might be the opposite .
    - **MoE-1 R3:** i never thought about being a dick .

    - **MoE-2 R1:** i guess there 's a reason for people to use their username too .
    - **MoE-2 R2:** i just use reddit , and the new design is pretty good .
    - **MoE-2 R3:** i see a lot of people using this as well as the same goes .

    - **MoE-3 R1:** you might also be able to get a hold of their different places with only their username .
    - **MoE-3 R2:** you might even be able to get a hold of them in different places with only their username .
    - **MoE-3 R3:** you might also be able to get a hold of my different places with the internet access .

    - **MoE-4 R1:** i 'm also loving the music so far .
    - **MoE-4 R2:** haha well i guess you are talking about the internet .
    - **MoE-4 R3:** haha good karma happens in a lot of us , but for me i 'm most on reddit .

A.7.2 CORNELL MOVIE CORPUS

1. **Sample Conversation:** (...) i guess i cut pretty loose in my day too .
    - **MoE-1 R1:** oh , i 'm fine .
    - **MoE-1 R2:** oh , yeah .
    - **MoE-1 R3:** oh , yeah .

    - **MoE-2 R1:** yes , darling . i 'm looking for you .
    - **MoE-2 R2:** nice . good .
    - **MoE-2 R3:** okay , we 'll see you .

    - **MoE-3 R1:** oh yeah . that 's nice .
    - **MoE-3 R2:** oh , it 's probably a good idea .
    - **MoE-3 R3:** oh , it 's worth it .

    - **MoE-4 R1:** it 's so beautiful . there 's a lot of money in there .
    - **MoE-4 R2:** i 'm so sure you 'll find out how that d go.
    - **MoE-4 R3:** i said it 's good call .

2. **Sample Conversation:** (...) i could get ben to tap into the baxter 's main power to generate enough voltage
    - **MoE-1 R1:** yes , sir .
    - **MoE-1 R2:** oh , yes , sir .
    - **MoE-1 R3:** oh , yes , sir .

    - **MoE-2 R1:** and i 'm sure it 's a good idea
    - **MoE-2 R2:** yes , ma'am . i 'm going to work .

- **MoE-2 R3:** oh , i 'm going to have a nice time .

- **MoE-3 R1:** hey . that 's the problem .
- **MoE-3 R2:** oh , that 's a possibility .
- **MoE-3 R3:** well . hope you enjoy it .

- **MoE-4 R1:** i do n't know . but i 'm sure it 's a very good idea .
- **MoE-4 R2:** i 'm glad you got your sense of humor back . think you 'll keep it ?
- **MoE-4 R3:** yes , that 's exactly what we got to do .

3. **Sample Conversation:** (...) or that he ca n't be bothered to write us a proper letter ? a dissertation on apples , we do n't need !
    - **MoE-1 R1:** no , sir .
    - **MoE-1 R2:** yes , sir .
    - **MoE-1 R3:** yes , sir .

    - **MoE-2 R1:** yes , sir . they 're telling me all this .
    - **MoE-2 R2:** he 's a bad guy , huh ?
    - **MoE-2 R3:** oh , that 's a good idea .

    - **MoE-3 R1:** oh . well , that 's good .
    - **MoE-3 R2:** oh , that sounds great .
    - **MoE-3 R3:** yes , he 's that sweetheart .

    - **MoE-4 R1:** he says he does n't feel good too . i thought you might know .
    - **MoE-4 R2:** yes . he was all right .
    - **MoE-4 R3:** jeez , right . he 's like a ghost !

A.8 Sample Utterances for EXP 2: Optimism Emotion Intent

A.8.1 Reddit Casual

1. **Sample Conversation:** (...) besides every anime intro and outro song ever at least from the series i like , i 've been getting a kick out of j rock in general .
    - **MoE-1 R1:** got a few more of my favorite songs !
    - **MoE-1 R2:** have you ever got a song that you can listen to ?
    - **MoE-1 R3:** i ve been listening to a lot of songs lately and i can relate .

    - **MoE-2 R1:** good lord , definitely send forever shit to the artist ...
    - **MoE-2 R2:** good for you , the fantastic video !
    - **MoE-2 R3:** good one , but the next one is , better !

    - **MoE-3 R1:** i got deep into idol music this year so i can relate a bit .
    - **MoE-3 R2:** i got big into pop music today so i can relate your success !
    - **MoE-3 R3:** i got deep into music recently so i will love my new year !

    - **MoE-4 R1:** good for you , it s a good anime intro haha
    - **MoE-4 R2:** might have some japanese jazz fusion and art
    - **MoE-4 R3:** good one , the anime and manga really fit me

2. **Sample Conversation:** (...) i got deep into idol music this year so i can relate a bit .
    - **MoE-1 R1:** what do you like about it ?
    - **MoE-1 R2:** what do you like to see ?
    - **MoE-1 R3:** i 'm not sure what to watch ?

- **MoE-2 R1:** good one , the next anime 's been awesome !,
- **MoE-2 R2:** good for you , my friend !
- **MoE-2 R3:** the best way to start the new year !

- **MoE-3 R1:** is it your favourite song ever ?
- **MoE-3 R2:** does it match your taste ?
- **MoE-3 R3:** is it really your favorite movie ?

- **MoE-4 R1:** good one , especially for the current age
- **MoE-4 R2:** thanks for your good song it gets more popular
- **MoE-4 R3:** good lord , the music is popular !

3. **Sample Conversation:** (...) some people have the same username across multiple websites such as steam or xbox .
    - **MoE-1 R1:** it 's not even a joke , but i think it is a scam .
    - **MoE-1 R2:** why people would use it as a joke ?
    - **MoE-1 R3:** how do you know how to deal with it ?

    - **MoE-2 R1:** what you can see is a new generation of the future.
    - **MoE-2 R2:** be aware of what their user names are and comments are made to post
    - **MoE-2 R3:** it depends on the mods and the subreddit , i see them as a possible community .

    - **MoE-3 R1:** you might even be able to get a hold of their different places with different people
    - **MoE-3 R2:** you might even be able to get a hold of their different places with different people than your own view .
    - **MoE-3 R3:** you may even be able to get a hold of your friend in different places with their username ?

    - **MoE-4 R1:** alright , so , especially reddit . might have to check out the real difference , especially with older ones !
    - **MoE-4 R2:** good on you for noticing when you mentioned that though not to be a problem !
    - **MoE-4 R3:** good for you , despite only being the mom and dad for me , i believe there are some old people now , still hoping to do that

A.8.2 CORNELL MOVIE CORPUS

1. **Sample Conversation:** (...) i guess i cut pretty loose in my day too .
    - **MoE-1 R1:** oh , i 'm fine .
    - **MoE-1 R2:** oh , yeah .
    - **MoE-1 R3:** oh , i 'm sorry .

    - **MoE-2 R1:** okay , take it easy .
    - **MoE-2 R2:** all right . everything 's been pretty .
    - **MoE-2 R3:** go ahead . the last time you 're right .

    - **MoE-3 R1:** i 'm talking about ...
    - **MoE-3 R2:** let 's see if i ever known .
    - **MoE-3 R3:** you learned from your american friends ?

    - **MoE-4 R1:** i'm sure it can be fine . it 's just you and me , brad .
    - **MoE-4 R2:** it 's so cool . really .
    - **MoE-4 R3:** you 're right . i 'm sure it will work . i could feel you in your position .

2. **Sample Conversation:** (...) i could get ben to tap into the baxter 's main power to generate enough voltage
    - **MoE-1 R1:** yes , sir . but we 're not going to do the same thing .
    - **MoE-1 R2:** no , sir .
    - **MoE-1 R3:** yes , sir .

    - **MoE-2 R1:** all right . it 's the way . i 'm going to get it .
    - **MoE-2 R2:** all right . you 'll do it .
    - **MoE-2 R3:** go ahead . i 'll be all right .

    - **MoE-3 R1:** it is ! i 'm stopping !
    - **MoE-3 R2:** that bastard ! i 'll gladly dispose of him in the name of the order , son of malkovich .
    - **MoE-3 R3:** i think they should stop and start looking right here i think they should stop and start looking right here .

    - **MoE-4 R1:** i 'm looking at it . i 'm going to get it right .
    - **MoE-4 R2:** there 's something else , major . i 'm starting to think what you said has a ton of beauty in it
    - **MoE-4 R3:** see , that 's what we 're here for .

3. **Sample Conversation:** (...) or that he ca n't be bothered to write us a proper letter ? a dissertation on apples , we do n't need !
    - **MoE-1 R1:** yes , sir . is that what you mean ?
    - **MoE-1 R2:** yes , sir .
    - **MoE-1 R3:** oh , yes , sir .

    - **MoE-2 R1:** all right . we 'll be able to figure out what 's wrong with him .
    - **MoE-2 R2:** okay , take it easy .
    - **MoE-2 R3:** all right . we 'll be able to figure out what 's going on .

    - **MoE-3 R1:** so ? the man got silver , china , rugs
    - **MoE-3 R2:** he 's out to the hospital ...
    - **MoE-3 R3:** i see , i was just thinking ... maybe it 's time to settle down .

    - **MoE-4 R1:** you do n't understand , i 've been in love with him !
    - **MoE-4 R2:** he 'll be back . i 'll wait for his car .
    - **MoE-4 R3:** i 'll think of something . i do n't think he 's a bad guy .

A.9 SAMPLE UTTERANCES FOR EXP 2: ANGER EMOTION INTENT

A.9.1 REDDIT CASUAL

1. **Sample Conversation:** (...) besides every anime intro and outro song ever at least from the series i like , i 've been getting a kick out of j rock in general .
    - **MoE-1 R1:** i got ta have this song , but i can relate to this .,
    - **MoE-1 R2:** i got this song a few months ago , and i can relate to a bit .
    - **MoE-1 R3:** i got my thoughts into this , but i can relate to this .

    - **MoE-2 R1:** why are you into anime ?,
    - **MoE-2 R2:** i lose interest in it .
    - **MoE-2 R3:** i 'm not an expert when it comes to it .

    - **MoE-3 R1:** i got deep into idol music this year so i can relate a bit .
    - **MoE-3 R2:** i got deep into this punk shows recently so i can relate with it .

- **MoE-3 R3:** i got ta learn rock this pop punk so i can relate with it .

- **MoE-4 R1:** keeping my speed for reading the anime shit .
- **MoE-4 R2:** why are you damn talking english only ?
- **MoE-4 R3:** and this is why i am not such a rich dude .

2. **Sample Conversation:** (...) i got deep into idol music this year so i can relate a bit .
   - **MoE-1 R1:** oh wow , that 's a good one .
   - **MoE-1 R2:** i like it , i 'm not a fan of the music genre , but i 've had a lot of music .
   - **MoE-1 R3:** oh wow .

   - **MoE-2 R1:** why are you considering the big ex ?
   - **MoE-2 R2:** i don 't watch anime in general at the moment .
   - **MoE-2 R3:** why are you considering anime good ?

   - **MoE-3 R1:** i can totally see that this was all about the songs i did n't talk to about .
   - **MoE-3 R2:** i would n't miss out on that video .
   - **MoE-3 R3:** i don t miss that so much .

   - **MoE-4 R1:** fuck yeah , i was there for like shit .
   - **MoE-4 R2:** why are you considering doing a lot of anime ?
   - **MoE-4 R3:** why do you think anime is bad ?

3. **Sample Conversation:** (...) some people have the same username across multiple websites such as steam or xbox .
   - **MoE-1 R1:** i think it all comes down to if they 're serious .
   - **MoE-1 R2:** i think it 's just how they are in their mid 20s how they are , they are used to their own lives .
   - **MoE-1 R3:** it might seem like a lot of people in their late 20s , they have a large problem with their accent .

   - **MoE-2 R1:** why do people assume they have their own identity ?
   - **MoE-2 R2:** it 's just a series of spam jokes and whatnot
   - **MoE-2 R3:** i think people are just dumb to the bottom line of their head .

   - **MoE-3 R1:** you might also be able to find a hold of them with different people than the previous ones .
   - **MoE-3 R2:** you might also be able to get a hold of them in different places with the username .
   - **MoE-3 R3:** you might even be able to get a hold of them in different places with their username .

   - **MoE-4 R1:** i dont know who downvoted you , but let people tell you the truth of your karma . why are you asking on reddit ?
   - **MoE-4 R2:** i dont necessarily agree with them but just became the first and its just became a good one .
   - **MoE-4 R3:** you 're around and just do n't use the internet for any specific reason , i think there are thousands of people watching .

A.9.2 Cornell Movie Corpus

1. **Sample Conversation:** (...) i guess i cut pretty loose in my day too .
   - **MoE-1 R1:** you do n't have to .
   - **MoE-1 R2:** no .
   - **MoE-1 R3:** you do n't have to have a good look .

- **MoE-2 R1:** i think you must be crazy . you ca n't tell me what you 're doing here !
- **MoE-2 R2:** you 're not the one trying to get away . who owns the gun ?
- **MoE-2 R3:** you do n't pull this cowboy shit , you 're gon na get him out ?

- **MoE-3 R1:** what do you know . i saw you ?
- **MoE-3 R2:** what the hell is that
- **MoE-3 R3:** oh , you do n't know . i 'm sorry .

- **MoE-4 R1:** what are you talking about ? you love your job .
- **MoE-4 R2:** you 're like an undercover guy !
- **MoE-4 R3:** you do n't get paid till we close tonight .

2. **Sample Conversation:** (...) i could get ben to tap into the baxter 's main power to generate enough voltage
   - **MoE-1 R1:** you do n't have to do this .
   - **MoE-1 R2:** you do n't know what you 're talking about .
   - **MoE-1 R3:** you do n't know what you 're talking about .

   - **MoE-2 R1:** you got to be kidding !
   - **MoE-2 R2:** we 're not going to do this .
   - **MoE-2 R3:** we want to frighten you away . so that you 'll move out

   - **MoE-3 R1:** do n't you dare , i 'll be all that ...
   - **MoE-3 R2:** so you 're telling the truth ?
   - **MoE-3 R3:** you think i 'm right

   - **MoE-4 R1:** you 're driving me nuts !
   - **MoE-4 R2:** you 're trying to kill me , are n't you ?
   - **MoE-4 R3:** you do n't get it , pham . the idea was we were ripping them off .

3. **Sample Conversation:** (...) or that he ca n't be bothered to write us a proper letter ? a dissertation on apples , we do n't need !
   - **MoE-1 R1:** do n't worry about it
   - **MoE-1 R2:** do n't you think he 's in the dark .
   - **MoE-1 R3:** do n't worry about it .

   - **MoE-2 R1:** i ca n't believe you did n't see that coming . you do n't want to lose this !
   - **MoE-2 R2:** what ? you 're not going to kill him .
   - **MoE-2 R3:** now what ? we 're going to have to do something

   - **MoE-3 R1:** because you do n't know what you are doing .
   - **MoE-3 R2:** oh , you do n't know . i guess it 's just not true .
   - **MoE-3 R3:** and what the hell is that ?

   - **MoE-4 R1:** he 'll kill you ! motherfucker ! !
   - **MoE-4 R2:** see , i told you he was gon na kill us all .
   - **MoE-4 R3:** and you do n't care what it says ?

A.10 Sample Utterances for EXP 2: Sadness Emotion Intent

A.10.1 Reddit Casual

1. **Sample Conversation:** (...) besides every anime intro and outro song ever at least from the series i like , i 've been getting a kick out of j rock in general .
   - **MoE-1 R1:** i got this feeling of thinking about this year and i can relate .

- **MoE-1 R2:** i got this feeling i can relate to this song a bit .
- **MoE-1 R3:** i got this song a lot , and i can relate to it .

- **MoE-2 R1:** i really miss the music i was listening to it .
- **MoE-2 R2:** i have n't heard of it for a long time .
- **MoE-2 R3:** i 'm not a fan of anime at all .

- **MoE-3 R1:** i got deep into idol music this year so i can relate a bit .
- **MoE-3 R2:** i got very deep into this show so i can relate with this most .
- **MoE-3 R3:** i got very into film school this year so i can relate with you .

- **MoE-4 R1:** no i would n't have watched anime or video .
- **MoE-4 R2:** i only watched it in japanese and i was at a time .
- **MoE-4 R3:** i feel like i m just listening to the same music again .

2. **Sample Conversation:** (...) i got deep into idol music this year so i can relate a bit .
   - **MoE-1 R1:** i 'm not the most interested in this movie .
   - **MoE-1 R2:** i 've been listening to the whole genre of the past .
   - **MoE-1 R3:** i 've been listening to a lot of people from the past few years .

   - **MoE-2 R1:** i really miss the music i 've listened to before
   - **MoE-2 R2:** i feel like i 've been listening to that music lately .
   - **MoE-2 R3:** i hate reddit and have no idea what it is

   - **MoE-3 R1:** i was also a huge fan of music so i do n't listen to them tv .
   - **MoE-3 R2:** i was very confused on my youtube was bc the record is bothering me .
   - **MoE-3 R3:** i 'm also very into film .

   - **MoE-4 R1:** no but i only listen to rap .
   - **MoE-4 R2:** i just watched anime and manga japan was probably the worst .
   - **MoE-4 R3:** i have n't heard anyone like anime .

3. **Sample Conversation:** (...) some people have the same username across multiple websites such as steam or xbox .
   - **MoE-1 R1:** i think it 's because people who use the internet , they do n't use them anymore .
   - **MoE-1 R2:** i think that 's the case with the people who have their own knowledge of their own .
   - **MoE-1 R3:** i think people are just too used to it in their 30s or at their own cult .

   - **MoE-2 R1:** i 've seen people get downvoted and it 's all weird .
   - **MoE-2 R2:** i even feel like it 's true for everyone .
   - **MoE-2 R3:** i 've heard people are n't on reddit .

   - **MoE-3 R1:** you might also be able to get a hold of their own place without similar type places .
   - **MoE-3 R2:** you might also be able to get a hold of them within the same price or not .
   - **MoE-3 R3:** you might also be able to get a hold of them within different countries or whatever they send you .

   - **MoE-4 R1:** i have noticed it on twitter , and facebook . i never have anonymous enough on twitter
   - **MoE-4 R2:** i have n't gotten around to it . i am literally wanting to delete my account .
   - **MoE-4 R3:** i do n't know why people are so dumb and useless .

A.11 SAMPLE UTTERANCES FOR EXP 3 (WITH FULL MOE DECISION FLOW)

A.11.1 REDDIT CASUAL, MOE-4 IQL

1. **Turn 0:** (...) when i actually look back at it , it 's not bad .
   - **MoE Primitive 1:** oh god those words mean a lot in general . Q**-value:** 0.15
   - **MoE Primitive 2:** when i get a good amount of things i put in the mood . Q**-value:** -0.32
   - **MoE Question 1:** i have n't listened to music , but i played a few songs , what 's it ? Q**-value:** -0.38
   - **MoE Question 2:** how much is it ? Q**-value:** -0.41
   - **MoE Exploration 1:** and that 's just me being the only way i can discover a life . Q**-value:** -0.26
   - **MoE Exploration 2:** i 'm 22 and have n't hugged a girl yet forget about kissing . Q**-value:** -0.55
   - **MoE Positive Sentiment 1:** thank you for the amazing reply , i definitely look at it to a high day ! Q**-value:** 0.45
   - **MoE Positive Sentiment 2:** good for you , i really love the genre so things i consider a best with as my life , it was a very difficult decision Q**-value:** 0.05
   - **MoE Negative Sentiment 1:** but on my own , i just do n't see the appeal anymore with how it is . Q**-value:** 0.05
   - **MoE Negative Sentiment 2:** i hate it when i 'm slow . Q**-value:** -0.46
   - **MoE Sentiment Coherence 1:** for me , i have a taste in music so i can not stand a lot of art or without it i would be so tired Q**-value:** -0.38
   - **MoE Sentiment Coherence 2:** the reason i love it , i have n't been in a while , but i feel like i was in a way that was when i was younger . Q**-value:** 0.25
   - **MoE Joy Emotion 1:** haha well i 'm going through a lot of shit . Q**-value:** -0.25
   - **MoE Joy Emotion 2:** haha yes i have my own one of my favorite things to do . Q**-value:** -0.62
   - **MoE Optimism Emotion 1:** good for you , i like the original music , it 's a new skill in painting , some of it , and just like every other Q**-value:** -0.14
   - **MoE Optimism Emotion 2:** good on you for doing the right thing , just remember that the past few months , i 've been putting it off , and have no idea about music Q**-value:** 0.69
   - **MoE Anger Emotion 1:** why do you think you re an adult ? Q**-value:** -0.16
   - **MoE Anger Emotion 2:** why are you not asking me a question ? Q**-value:** -0.29
   - **MoE Sadness Emotion 1:** i do n't have much of an effect on my mood , what should i do ! Q**-value:** 0.60
   - **MoE Sadness Emotion 2:** i have n't done that yet but i 'm going to check out . Q**-value:** -0.25

   - **Final Bot Response:** Q**-value:** good on you for doing the right thing , just remember that the past few months , i 've been putting it off and have no idea about music Q**-value:** 0.69

   - **GPT-based Next-user Response:** just love these memory of having that moment of myself and lots to remember the past few years on life of music .

2. **Turn 1:** (...) just love these memory of having that moment of myself and lots to remember the past few years on life of music .
   - **MoE Primitive 1:** it has become an actual meme , and then on a cracked up . Q**-value:** 0.21
   - **MoE Primitive 2:** i just love music , what genre would it be ? Q**-value:** 0.42
   - **MoE Question 1:** how much do you enjoy now ? Q**-value:** -0.15
   - **MoE Question 2:** how do you discover new music ? Q**-value:** 0.08
   - **MoE Exploration 1:** but i was thinking of going to be a one , and maybe <unk> my favorite shows . Q**-value:** 0.35

- **MoE Exploration 2:** and it 's a good time to focus on something else , like a good <unk> ' , <unk> out of the good things . Q**-value:** -0.43
- **MoE Positive Sentiment 1:** thanks for the advice , i really hope things turn to someone who does n't have a life and does n't deserve it and as much to be Q**-value:** -0.51
- **MoE Positive Sentiment 2:** thank you for the inspiration ! Q**-value:** 0.13
- **MoE Negative Sentiment 1:** but i hate when i have to do it all . Q**-value:** 0.18
- **MoE Negative Sentiment 2:** i am not one of those and i do n't know about to be honest . Q**-value:** -0.47
- **MoE Sentiment Coherence 1:** for the future , i love it and remember that song <unk> would be a very nice place to achieve something like <unk> or goals . Q**-value:** -0.57
- **MoE Sentiment Coherence 2:** for me , i love the feeling of wanting to have a good christmas tree since i was a kid Q**-value:** -0.77
- **MoE Joy Emotion 1:** haha , i am enjoying a lot of it , so i personally even know nothing about it . Q**-value:** 0.09
- **MoE Joy Emotion 2:** haha , this is one of my favorite things to listen in the genre . Q**-value:** 0.71
- **MoE Optimism Emotion 1:** good for you , the best of luck in the future ! Q**-value:** 0.13
- **MoE Optimism Emotion 2:** for me , the only thing i really like , is the music i like , and the way it gives me time to myself and be Q**-value:** 0.48
- **MoE Anger Emotion 1:** and what is an example of things you could do ? Q**-value:** -0.11
- **MoE Anger Emotion 2:** and i do n't like music , i 've felt so far . Q**-value:** -0.02
- **MoE Sadness Emotion 1:** i do n't know what to do with my knowledge but i have no idea what it is . Q**-value:** -0.03
- **MoE Sadness Emotion 2:** i have no clue why i do that these days . Q**-value:** -0.31

- **Final Bot Response:** haha , this is one of my favorite things to listen in the genre . Q**-value:** 0.71

- **GPT-based Next-user Response:** I wonder how how it 'd go . but i do enjoy them . they feel live and have the meaning of life .

A.11.2 Reddit Casual, MoE-4 KLC

1. **Turn 0:** (...) when i actually look back at it , it 's not bad .
   - **MoE Primitive 1:** well i 'm glad you 're feeling better . Q**-value:** 0.38
   - **MoE Primitive 2:** did you get a song that you 're listening to ? Q**-value:** -0.06
   - **MoE Question 1:** what 's it worth ? Q**-value:** 0.35
   - **MoE Question 2:** how did you get out of it ? Q**-value:** 0.12
   - **MoE Exploration 1:** lately i 've been a lot of good things to do . Q**-value:** -0.55
   - **MoE Exploration 2:** i love this song , and i will be happy with the music . Q**-value:** -0.26
   - **MoE Positive Sentiment 1:** i 'm glad you 're able to find a great life ! Q**-value:** 0.40
   - **MoE Positive Sentiment 2:** i hope you have a great day ! Q**-value:** 0.39
   - **MoE Negative Sentiment 1:** i do n't know if it 's worth a listen to it . Q**-value:** 0.22
   - **MoE Negative Sentiment 2:** i do n't know if i 'm going to be a bit stressful . Q**-value:** 0.17
   - **MoE Sentiment Coherence 1:** i 'll be able to do something with my own , but i 'll be able to find a new genre . Q**-value:** -0.11
   - **MoE Sentiment Coherence 2:** i might try to find some new songs that i 'm not sure if i 'm a bit interested in the music . Q**-value:** 0.20
   - **MoE Joy Emotion 1:** i feel so much better now Q**-value:** -0.45
   - **MoE Joy Emotion 2:** i want to do that myself . Q**-value:** -0.22
   - **MoE Optimism Emotion 1:** be proud of yourself and happy for the first time i 've been able to make a great life . Q**-value:** 0.29

- **MoE Optimism Emotion 2:** be proud of yourself and be happy for the first time i 've been to a great life . Q**-value:** 0.09
- **MoE Anger Emotion 1:** it 's so much easier to say than done ! Q**-value:** -0.50
- **MoE Anger Emotion 2:** i need to start doing something productive , but i 'm not doing anything . Q**-value:** -0.11
- **MoE Sadness Emotion 1:** i just started a year ago and i 'm not sure what to do . Q**-value:** 0.16
- **MoE Sadness Emotion 2:** i 've been trying to find some new songs that are n't a thing . Q**-value:** 0.12

- **Final Bot Response:** i 'm glad you 're able to find a great life ! Q**-value:** 0.40

- **GPT-based Next-user Response:** yes , life feels young and happy again .

2. **Turn 1:** (...) yes , life feels young and happy again .

- **MoE Primitive 1:** i feel the same way . Q**-value:** 0.061
- **MoE Primitive 2:** i 've been through a lot of pain , but i do n't know how to do it . Q**-value:** -0.045
- **MoE Question 1:** how do you feel about it ? Q**-value:** -0.35
- **MoE Question 2:** i can understand that . how do you cope with this ? Q**-value:** 0.03
- **MoE Exploration 1:** but that 's how i feel about it and i 'm not going to be able to enjoy it . Q**-value:** 0.15
- **MoE Exploration 2:** because it 's not a good feeling to be able to do something i Q**-value:** -0.56
- **MoE Positive Sentiment 1:** be proud of yourself and be the best part of life , i 'm happy that you 're happy . Q**-value:** 0.85
- **MoE Positive Sentiment 2:** i hope i can find a great life too ! Q**-value:** 0.65
- **MoE Negative Sentiment 1:** i do n't know if it 's going to be the problem . Q**-value:** -0.22
- **MoE Negative Sentiment 2:** yeah i feel like i have a hard time . Q**-value:** 0.18
- **MoE Sentiment Coherence 1:** i might try to get into the good stuff i 'm not going to do . Q**-value:** 0.05
- **MoE Sentiment Coherence 2:** i 'll be able to find the better , but i 'm not sure if i can . Q**-value:** -0.19
- **MoE Joy Emotion 1:** i just enjoy playing music , and i 'm happy for myself . Q**-value:** 0.55
- **MoE Joy Emotion 2:** just in the mood , i 'm going to enjoy a bit of it . Q**-value:** -0.33
- **MoE Optimism Emotion 1:** just in the mood , i 'm going to enjoy a bit of it . Q**-value:** -0.61
- **MoE Optimism Emotion 2:** i just enjoy playing music , and i 'm happy for myself . Q**-value:** 0.28
- **MoE Anger Emotion 1:** i just started a new year ago and i 'm not a huge fan of the music . Q**-value:** -0.43
- **MoE Anger Emotion 2:** why do you have to do this ? Q**-value:** -0.59
- **MoE Sadness Emotion 1:** got any hobbies that i feel like i 'm alone . Q**-value:** -0.13
- **MoE Sadness Emotion 2:** do you have any advice or am i just a little curious about this ? Q**-value:** -0.08

- **Final Bot Response:** be proud of yourself and be the best part of life , i 'm happy that you 're happy .

- **GPT-based Next-user Response:** thank you i 'd proud of you 've been enjoy life too !

A.11.3 Reddit Casual, MoE-4 Ensemble-Q

1. **Turn 0:** (...) when i actually look back at it , it ’s not bad .
    - **MoE Primitive 1:** oh god those words mean a lot in general too . **Q-value:** 0.03
    - **MoE Primitive 2:** when i m in a really bad mood , i ll have a look about myself . **Q-value:** -0.05
    - **MoE Question 1:** how do you know what ’s good ? **Q-value:** -0.25
    - **MoE Question 2:** how in the world do you have all the time ? **Q-value:** -0.35
    - **MoE Exploration 1:** and that ’s what i think of when i want to have a good idea . **Q-value:** 0.52
    - **MoE Exploration 2:** at least i have a couple of hours of listening to the same tunes . **Q-value:** 0.28
    - **MoE Positive Sentiment 1:** thanks for the good things , i ’ve found in 2018 , still , and if you like it ’s a little bit of inspiration , my own memories . **Q-value:** 0.16
    - **MoE Positive Sentiment 2:** thank you for your recommendations , i definitely do n’t have a fantastic artist , it ’s absolutely amazing to me , but i really want to learn more **Q-value:** -0.4
    - **MoE Negative Sentiment 1:** but on my mind , i ca n’t even focus on the music . **Q-value:** -0.06
    - **MoE Negative Sentiment 2:** but no i just do n’t know ... **Q-value:** 0.55
    - **MoE Sentiment Coherence 1:** thanks for the recommendation i really enjoyed it ! **Q-value:** -0.04
    - **MoE Sentiment Coherence 2:** that ’s what i ’ve heard about in the first place . **Q-value:** -0.13
    - **MoE Joy Emotion 1:** not bad , i have all my <unk> on the side and <unk> . **Q-value:** -0.18
    - **MoE Joy Emotion 2:** good for you , for me , my best friend is <unk> **Q-value:** -0.27
    - **MoE Optimism Emotion 1:** good on you for the time , but it ’s actually a bit of a hiatus from the genre your get to your development aspect , regardless of songs **Q-value:** -0.67
    - **MoE Optimism Emotion 2:** you know what i m missing out on . **Q-value:** 0.24
    - **MoE Anger Emotion 1:** i do n’t like the term toxic masculinity , but i think this idea of pride is accurately described . **Q-value:** -0.36
    - **MoE Anger Emotion 2:** why do you think the weed smell is a good idea ? **Q-value:** -0.64
    - **MoE Sadness Emotion 1:** i do n’t have much of a high expectation in my life . **Q-value:** 0.15
    - **MoE Sadness Emotion 2:** i ’m no expert on turtles and different types of turtle species , i just saw an interesting documentary on sea turtles . **Q-value:** -0.04

    - **Final Bot Response:** but no i just do n’t know ... **Q-value:** 0.55

    - **GPT-based Next-user Response:** i agree , just make sure what you want to do and figure out how to proud
2. **Turn 1:** (...) i agree , just make sure what you want to do and figure out how to proud
    - **MoE Primitive 1:** it has become an actual effect on my mood . **Q-value:** 0.65
    - **MoE Primitive 2:** but i ’ve discovered that by not giving any fucks about what i want . **Q-value:** 0.13
    - **MoE Question 1:** how the fuck do you think about it ? **Q-value:** -0.06
    - **MoE Question 2:** how are you enjoying life , what you used to do ? **Q-value:** 0.18
    - **MoE Exploration 1:** that s why i have a favourite meme playlist right . **Q-value:** -0.24
    - **MoE Exploration 2:** for me , i do n’t know what a genre is hit these days . **Q-value:** -0.34
    - **MoE Positive Sentiment 1:** for me , i love the good drawing and trying to keep it up , but it s not a big deal , i like a lot of it **Q-value:** 0.23
    - **MoE Positive Sentiment 2:** good for you , for me , my best friend is music **Q-value:** 0.12

- **MoE Negative Sentiment 1:** i hate to admit it , i just look through the whole world . Q**-value:** 0.08
- **MoE Negative Sentiment 2:** i hate to admit it , i do n't know what i like to listen Q**-value:** 0.28
- **MoE Sentiment Coherence 1:** taste is all subjective and i have no matter to taste or texture other things on the genre . Q**-value:** -0.55
- **MoE Sentiment Coherence 2:** although i 'm on youtube , i enjoy old music . Q**-value:** 0.17
- **MoE Joy Emotion 1:** good for you , there s really no way to do it , but i m in the mood for it , in my day Q**-value:** -0.23
- **MoE Joy Emotion 2:** good on you for the time , i think it 's a little complicated haha Q**-value:** 0.45
- **MoE Optimism Emotion 1:** good one , i have an amazing time , but , the little it can be so challenging , especially to discover the passion . Q**-value:** 0.75
- **MoE Optimism Emotion 2:** this is the best recommendation ever , and at a good old dance moves are n't any other , or at the same time Q**-value:** 0.28
- **MoE Anger Emotion 1:** no taste in music , but the more <unk> ' is my favorite song ever on things . Q**-value:** -0.48
- **MoE Anger Emotion 2:** also , i have n't watched an entire year of it . Q**-value:** -0.08
- **MoE Sadness Emotion 1:** why do i have to not use it ? Q**-value:** -0.0
- **MoE Sadness Emotion 2:** i do n't know why i stopped putting on so much . Q**-value:** 0.15

- **Final Bot Response:** good one , i have an amazing time , but , the little it can be so challenging , especially to discover the passion . Q**-value:** 0.75

- **GPT-based Next-user Response:** i already have a really vivid of memories of those conversations , thats what i have heard from .

A.11.4 Cornell Movie Corpus, MoE-2 IQL

1. **Turn 0:** (...) excuse me
   - **MoE Primitive 1:** you , huh ? is that you ? Q**-value:** 0.17
   - **MoE Primitive 2:** i 'll promise you . but i 'd not like one of them , do n't worry . Q**-value:** 0.28
   - **MoE Question 1:** what ? what ? because he was worried about ? Q**-value:** -0.34
   - **MoE Question 2:** what ? oh , i thought maybe you did too ? Q**-value:** -0.28
   - **MoE Exploration 1:** he asked you for a few bucks ? i want him along , anyway Q**-value:** 0.05
   - **MoE Exploration 2:** i do n't know . i 've been hearing from you since i left . Q**-value:** -0.01
   - **MoE Positive Sentiment 1:** he says you 're very nice . he has a gun . Q**-value:** -0.46
   - **MoE Positive Sentiment 2:** that 's right . he said that his name was familiar and the formal negotiation is handled ' Q**-value:** 0.18
   - **MoE Negative Sentiment 1:** i do n't know . i packed up my shit and ran away to here . Q**-value:** -0.22
   - **MoE Negative Sentiment 2:** no , i was n't gon na see you tonight . Q**-value:** -0.59
   - **MoE Sentiment Coherence 1:** i do n't know what happened . i 've never been there before . Q**-value:** 0.05
   - **MoE Sentiment Coherence 2:** i do n't know . i was with him and you 're all i know is that these years i thought you were going to do it Q**-value:** -0.35
   - **MoE Joy Emotion 1:** i 'm glad you got your memory back , father ? Q**-value:** 0.33
   - **MoE Joy Emotion 2:** haha , it is all good . Q**-value:** -0.17
   - **MoE Optimism Emotion 1:** i 'm so sure you do . and he 's got lots of friends left Q**-value:** -0.07
   - **MoE Optimism Emotion 2:** i was n't saying anything . i was singing . Q**-value:** 0.16

- **MoE Anger Emotion 1:** What ? you do n't pull that ? Q**-value:** -0.77
- **MoE Anger Emotion 2:** you did n't work for him , did n't you ? Q**-value:** -0.56
- **MoE Sadness Emotion 1:** he does n't like me ? Q**-value:** -0.15
- **MoE Sadness Emotion 2:** i 'm sorry , i was n't feeling that way . Q**-value:** 0.21

- **Final Bot Response:** i 'm glad you got your memory back , father ? Q**-value:** 0.33

- **GPT-based Next-user Response:** sure yea . you see i should worry nothing and do exactly as said

2. **Turn 1:** (...) sure yea . you see i should worry nothing and do exactly as said
   - **MoE Primitive 1:** i went down to meet the president . Q**-value:** -0.14
   - **MoE Primitive 2:** he 's the one i told him you . Q**-value:** 0.08
   - **MoE Question 1:** what ? what did you think he wanted to do , then ? Q**-value:** -0.35
   - **MoE Question 2:** what did you think ? when did you talk about it ? Q**-value:** -0.05
   - **MoE Exploration 1:** i 'm a little girl . if you do n't get me a girlfriend , maybe five bucks . Q**-value:** -0.20
   - **MoE Exploration 2:** it was like tommy was possessed or something . there was a hate in his eyes that i could n't really call human . Q**-value:** -0.41
   - **MoE Positive Sentiment 1:** it 's so nice to see you again . Q**-value:** 0.45
   - **MoE Positive Sentiment 2:** it was so cool . you know who 's being a jerk . Q**-value:** 0.05
   - **MoE Negative Sentiment 1:** i do n't want to be a fool . you do n't want to look ? Q**-value:** -0.04
   - **MoE Negative Sentiment 2:** i do n't care what you say ! i do n't want to fight you . kinda <unk> . Q**-value:** -0.46
   - **MoE Sentiment Coherence 1:** no . i think he 's in the house . Q**-value:** -0.43
   - **MoE Sentiment Coherence 2:** i do n't know what else you 'd call it . Q**-value:** -0.35
   - **MoE Joy Emotion 1:** it 's not that bad , it just the way you always wanted it to be , or did you ? Q**-value:** 0.33
   - **MoE Joy Emotion 2:** there you go . i like you . Q**-value:** 0.34
   - **MoE Optimism Emotion 1:** i 'm sure he 's very flattering , birdie , and i 'm sure there 's nothing wrong with that . Q**-value:** -0.25
   - **MoE Optimism Emotion 2:** he said it 's up to us . Q**-value:** 0.02
   - **MoE Anger Emotion 1:** you do n't get paid till you calm down . you 're sure you want to stay ? Q**-value:** -0.26
   - **MoE Anger Emotion 2:** you say that because you have no idea ? Q**-value:** -0.31
   - **MoE Sadness Emotion 1:** no , i 'm surprised what three years ago . i do n't know . Q**-value:** -0.05
   - **MoE Sadness Emotion 2:** i 'm sorry , i was n't feeling like this , all . Q**-value:** -0.10

   - **Final Bot Response:** it 's so nice to see you again . Q**-value:** 0.45

   - **GPT-based Next-user Response:** we think about you too , love you ,

A.11.5 CORNELL MOVIE CORPUS, MOE-2 KLC

1. **Turn 0:** (...) excuse me
   - **MoE Primitive 1:** you , huh ? is that you ? Q**-value:** -0.42
   - **MoE Primitive 2:** jesus , not sure . how are you ? Q**-value:** -0.24
   - **MoE Question 1:** what ? what ? you think he wanted to kill the baby ? Q**-value:** -0.091
   - **MoE Question 2:** what 's the matter ? Q**-value:** -0.080
   - **MoE Exploration 1:** i 'll destroy you , i swear it . Q**-value:** -0.51
   - **MoE Exploration 2:** he asked you for a few bucks ? do you want him along ? Q**-value:** -0.13

- **MoE Positive Sentiment 1:** sure , i 'd love to see you again . Q**-value:** -0.07
- **MoE Positive Sentiment 2:** it 's so beautiful . he was crazy about you . Q**-value:** 0.37
- **MoE Negative Sentiment 1:** i do n't want to hear it . i know you 're talking to him . Q**-value:** -0.11
- **MoE Negative Sentiment 2:** i did n't want to get involved . did you ever think to work with the police station chiefs . Q**-value:** -0.24
- **MoE Sentiment Coherence 1:** i do n't know what else you 'd call it . Q**-value:** -0.26
- **MoE Sentiment Coherence 2:** i was tired , what is that for ? Q**-value:** -0.26
- **MoE Joy Emotion 1:** he 's likely to be in love with you . Q**-value:** -0.24
- **MoE Joy Emotion 2:** i 'll be right there . i 'll have the money , you got that go . Q**-value:** 0.25
- **MoE Optimism Emotion 1:** i 'm sure you 'll remember that , in the first date ? Q**-value:** 0.25
- **MoE Optimism Emotion 2:** i 'm surprised you do n't know . when i was a kid , i live off the street . Q**-value:** 0.12
- **MoE Anger Emotion 1:** you do n't sound like it ! Q**-value:** -0.61
- **MoE Anger Emotion 2:** you were n't attracted to him ? Q**-value:** -0.11
- **MoE Sadness Emotion 1:** i 'm sorry , i do n't know how to make it out . Q**-value:** -0.17
- **MoE Sadness Emotion 2:** i 'm sorry , i do n't know what to say . Q**-value:** 0.05

- **Final Bot Response:** it 's so beautiful . he was crazy about you . Q**-value:** 0.37

- **GPT-based Next-user Response:** i 'm glad to know . but tonight mr. wade may not come here .

2. **Turn 1:** (...) i 'm glad to know . but tonight mr. mel may not come here .
    - **MoE Primitive 1:** certainly , when i get to go up to your room . Q**-value:** -0.33
    - **MoE Primitive 2:** no . do n't even be serious . but i 'm telling me the truth . Q**-value:** -0.21
    - **MoE Question 1:** what do you think he got to lose ? Q**-value:** -0.06
    - **MoE Question 2:** then what 's he talking about ? i 'm sorry . Q**-value:** -0.10
    - **MoE Exploration 1:** i just want to hear the details ... Q**-value:** 0.39
    - **MoE Exploration 2:** i do n't want to hear you anymore . you made love ? Q**-value:** -0.09
    - **MoE Positive Sentiment 1:** yes . i can live here with you Q**-value:** 0.25
    - **MoE Positive Sentiment 2:** right , youth . that 's why i need you to lead the troops . Q**-value:** -0.45
    - **MoE Negative Sentiment 1:** i do n't know . he fell for you . Q**-value:** -0.35
    - **MoE Negative Sentiment 2:** i do n't blame him . what happened ? Q**-value:** 0.01
    - **MoE Sentiment Coherence 1:** but you did n't think about it . what 's the one said . Q**-value:** -0.08
    - **MoE Sentiment Coherence 2:** i do n't know . i thought i might . but i 'm sure he 's quite human . Q**-value:** -0.06
    - **MoE Joy Emotion 1:** it 's just you know , i mean . it 's been a pleasure . Q**-value:**
    - **MoE Joy Emotion 2:** i 'm sure he 's the right person . Q**-value:**
    - **MoE Optimism Emotion 1:** yeah , i know . it 's been so long . Q**-value:** -0.18
    - **MoE Optimism Emotion 2:** i 'm sure you have n't done anything . Q**-value:** -0.22
    - **MoE Anger Emotion 1:** what ? i do n't know what else ? Q**-value:** 0.13
    - **MoE Anger Emotion 2:** you do n't have to prove it ? Q**-value:**
    - **MoE Sadness Emotion 1:** i was n't aware of that . but this is the beauty of it . Q**-value:** 0.55
    - **MoE Sadness Emotion 2:** clark . i would n't want to come in tonight . Q**-value:** -0.24

    - **Final Bot Response:** i was n't aware of that . but this is the beauty of it . Q**-value:** 0.55

- **GPT-based Next-user Response:** i 'm surprised you do n't know . he said some good thing about you too

A.11.6 CORNELL MOVIE CORPUS, MOE-2 ENSEMBLE-Q

1. **Turn 0:** (...) excuse me
    - **MoE Primitive 1:** i 'll promise you . but i do n't have any new evidence . Q**-value:** -0.19
    - **MoE Primitive 2:** no , it 's not the same thing . you do n't have the same . Q**-value:** -0.26
    - **MoE Question 1:** what ? what do you think happened here ? Q**-value:** 0.45
    - **MoE Question 2:** what did you think ? i 'm going to tell you what . it 's over , boy . Q**-value:** 0.11
    - **MoE Exploration 1:** it 's a lot of money for you to do . you know , i mean , you know , i 'm not going to do that Q**-value:** 0.04
    - **MoE Exploration 2:** it 's the only way i can do it . i 'm not going to be a writer . Q**-value:** 0.05
    - **MoE Positive Sentiment 1:** yes . it 's very nice . Q**-value:** -0.20
    - **MoE Positive Sentiment 2:** yes , it 's true . Q**-value:** -0.32
    - **MoE Negative Sentiment 1:** i do n't want to be here . i 'm not in love with you . Q**-value:** -0.25
    - **MoE Negative Sentiment 2:** do you want to be sure ? i want to know what you want . Q**-value:** -0.16
    - **MoE Sentiment Coherence 1:** i do n't know what you 're talking about . Q**-value:** -0.15
    - **MoE Sentiment Coherence 2:** you do n't have to be a cop . you 're not going to be there . Q**-value:** 0.03
    - **MoE Joy Emotion 1:** i 'm so sure you could have been anything ? Q**-value:** -0.28
    - **MoE Joy Emotion 2:** it was you , was n't it ? i 've seen your show . Q**-value:** 0.11
    - **MoE Optimism Emotion 1:** it 's not your fault . Q**-value:** 0.15
    - **MoE Optimism Emotion 2:** that 's all right , you 're gon na be fine . Q**-value:** -0.51
    - **MoE Anger Emotion 1:** you do n't pull that kick out your ass , you 'll try ? Q**-value:** -0.26
    - **MoE Anger Emotion 2:** you do n't have to do this to me ! Q**-value:** -0.53
    - **MoE Sadness Emotion 1:** i 'm sorry , i did n't know it . all i know is he did n't get to be here first . Q**-value:** 0.23
    - **MoE Sadness Emotion 2:** i 'm sorry to trouble you , but i have to . Q**-value:** 0.15

    - **Final Bot Response:** what ? what do you think happened here ? Q**-value:** 0.45

    - **GPT-based Next-user Response:** jesus , not sure . huh ? is that you ?

2. **Turn 1:** (...) jesus , not sure . huh ? is that you ?
    - **MoE Primitive 1:** we ran up against a platoon of soldiers . there were actually walking dead . Q**-value:** -0.05
    - **MoE Primitive 2:** no . it 's him , but he did n't make the choice . Q**-value:** 0.87
    - **MoE Question 1:** what ? what would you think ? i 'm going to be here . i 've never seen this guy . Q**-value:** 0.75
    - **MoE Question 2:** what did you think ? when did you talk about it ? Q**-value:** 0.18
    - **MoE Exploration 1:** that 's all i ever wanted . a man with the world without a different girl from one , desert one 's a dead , and later . Q**-value:** -0.51
    - **MoE Exploration 2:** i dunno ... police . Q**-value:** -0.02
    - **MoE Positive Sentiment 1:** it was like , was n't it ? you 're far too good . Q**-value:** -0.44
    - **MoE Positive Sentiment 2:** it 's so nice to see you again . Q**-value:** -0.56

- **MoE Negative Sentiment 1:** oh , i do n't care . but it 's got to be different . Q**-value:** -0.11
- **MoE Negative Sentiment 2:** i do n't want to hurt you . he 's not trying to get involved . Q**-value:** -0.30
- **MoE Sentiment Coherence 1:** no . i was just going to say , you know . Q**-value:** -0.45
- **MoE Sentiment Coherence 2:** i do n't know what else you 'd call it . Q**-value:** -0.53
- **MoE Joy Emotion 1:** i can be sure of that when you want your opinion . Q**-value:** -0.07
- **MoE Joy Emotion 2:** it was so obvious . i mean , all of nicky 's half assed mechanics , they were real signal happy . Q**-value:** 0.28
- **MoE Optimism Emotion 1:** i 'm sure he 's very flattering , birdie , and i 'm sure there 's nothing wrong with that . Q**-value:** -0.15
- **MoE Optimism Emotion 2:** there you go . that 's the truth . Q**-value:** 0.08
- **MoE Anger Emotion 1:** you do n't like him , do you ? Q**-value:** -0.07
- **MoE Anger Emotion 2:** he 's sleeping in the back , does n't he ? Q**-value:** -0.31
- **MoE Sadness Emotion 1:** no i was right behind you , but you 're too dumb to turn around . Q**-value:** 0.07
- **MoE Sadness Emotion 2:** i do n't know , i 've seen him . that 's the name of a man i thought . Q**-value:** -0.25

- **Final Bot Response:** no . it 's him , but he did n't make the choice . Q**-value:** 0.87

- **GPT-based Next-user Response:** that left me a surprise . did he say something ?

B EXPERIMENTAL DETAILS

This section describes more details about our experimental setup to evaluate the algorithms.

B.1 KL-CONTRASTIVE CONSTRAINT

Recall that the main purpose of the KL-constraint is to enforce consistency in the latent variables predicted by the semantic generator and the posterior. However, since both $\mathcal{G}_0(z'|z)$ and $\rho(z'|z, Y)$ are both models to be learned, if the data can be modeled with a single, stationary LM that does not depend on the latent space, one trivial degenerated solution (that satisfies the KL constraint) is to have both of these models output very small values, which would impede the generalizability (and diversity) of the LM embedding, especially to unseen conversations. Utilizing analogous arguments as in Shu et al. (2020); den Oord et al. (2018) that connects KL distribution matching and mutual information maximization in representation learning, we tackle this issue by considering the *KL-contrastive* constraint, i.e., let $F(z, Y) := \mathrm{KL}_{z'}\left(\rho(z'|z, Y) \mid\mid \mathcal{G}_0(z'|z)\right)$, we replace the constraint in the optimization problem in (1) with

$$\widehat{\mathbb{E}}_{z=\Phi(\mathbf{X}),Y\sim D}\left[F(z, Y) + \alpha \cdot \log \frac{\exp(-F(z, Y))}{\int_{Y'} \exp(-F(z, Y))}\right] \leq \epsilon_{\mathrm{KL}},$$

where $\alpha > 0$ is a trade-off factor in the KL-constrastive constraint. While the first part of the constraint limits $F(z, Y)$, the second part enforces $F(z, Y)$ to a much higher value in the positive samples (w.r.t. ground-truth next utterance) than in the negative samples (w.r.t. other next-utterance candidates). Therefore, this constraint can prevent the aforementioned degeneration issue and inject flexibility to control the size of the latent representation space.

B.2 DISCRETIZATION OF Φ AND $\mathcal{G}_i$

Inspired by various works which use discrete bottlenecks (Razavi et al., 2019; Hafner et al., 2020; Yang and Nachum, 2021), we employ discretization of Φ and $\mathcal{G}_0$ to encourage better generalization to out-of-distribution inputs. Specifically, we parameterize $\Phi(\mathbf{X}) = \mathrm{disc}_{16}(\widetilde{\Phi}(\mathbf{X}))$, where the output of $\widetilde{\Phi}$ is a continuous-valued vector in $\mathbb{R}^d$ and the discretization operator disc_K works as follows:

- Given vector $v \in \mathbb{R}^d$, split v into $v_1, \ldots, v_{d/K}$ vectors, each of size K.
- For each v_i, sample a one-hot vector v'_i of size K based on a softmax categorical distribution with logits given by v_i.
- Concatenate the vectors v'_i to yield a multi-hot vector v' of size d.
- When computing the backward pass on v', use straight-through gradients; i.e., $\frac{\partial v'_i}{\partial v_i} = \frac{\partial \mathrm{softmax}(v_i)}{\partial v_i}$.

We similarly discretize z', the output of $\mathcal{G}_0$. Namely, we first parameterize $\mathcal{G}_i$ as a Gaussian and then discretize that to form a multinomial distribution that produces multi-hot (d/K-hot, $K = 16$) vectors. We employ the same discretization for the experts $\mathcal{G}_i$ and the prior $\rho(z'|z, Y)$.

To avoid convergence to local minima for Φ, we employ an entropy regularizer on the discrete distribution it is sampled from. Namely, we use an adaptive entropy regularizer in the style of (Haarnoja et al., 2018) with a fixed target entropy that is a hyper-parameter.

B.3 MODEL PARAMETERS

In this section, the model parameters are described for MoE-1, 2, 3 and 4. All of these models represent Mixture of Experts and are based on transformer (Vaswani et al., 2017). Transformer is an encoder-decoder based model that uses self-attention to capture relationships between the elements of the sequence. In our implementation, we used multi-head attention with implementation similar to https://www.tensorflow.org/text/tutorials/transformer#point_wise_feed_forward_network.

MoE-1 and MoE-2 use the simple transformer architecture, while MoE-3 and MoE-4 use the same encoder architecture as BERT (Devlin et al., 2018). MoE-1 uses much smaller latent distribution models $\{\mathcal{G}_i\}$ than MoE-2; MoE-3 uses the pre-trained BERT encoder https://tfhub.dev/

tensorflow/bert_en_uncased_L-12_H-768_A-12/4, while MoE-4 trains that from scratch.

The transformer model parameters for the simple transformer architecture are summarized in Table 11:

Parameter	**Value**
Number of layers	2
Embedding hidden size	256
FFN inner hidden size	512
Attention heads	8
Key size	256
Value size	256
Dropout	0.1

Table 11: Simple Transformer Architecture

The BERT based transformer model is similar to the architecture from Table 11. The differences for a pre-trained BERT model are captured in Table 12.

Parameter	**Value**
Embedding hidden size	768
Number of layers	12
Attention heads	12

Table 12: BERT-based Transformer Architecture

The differences for a BERT model we train from scratch are captured in Table 13.

Parameter	**Value**
Embedding hidden size	768
Number of layers	2
Attention heads	8

Table 13: Trainable, Smaller "BERT"-based Transformer Architecture

Latent distributions $\{\mathcal{G}_i\}$ are implemented as FFN that model mean and variance of the normal distribution. The MoE-1, 2, 3, 4 use different values for the hidden size of the neural network. Additionally, MoE-1 and 3 use target entropy of 0.1, MoE-2 and 4 use target entropy of 1.0. The common parameters for FFN are captured in Table 14 (note: FFN has a final layer without an activation).

Finally, Table 15 shows individual parameters distinct for each of MoE models.

B.4 Computational resources

Training and evaluation were run on 8 GPU instances with 20GB of RAM and a NVIDIA Tesla P100 graphics card.

B.5 Dataset

We trained our models on Reddit Casual and Cornell Movie conversational datasets. Both datasets were downloaded from Neural Chat datasets of the MIT Media Lab https://affect.media.mit.edu/neural_chat/datasets. They contain conversational exchanges between pairs of speakers. Each batch of training data contains a subset of such conversations. The Reddit Casual is about 3 times bigger than Cornell corpus.

$\{\mathcal{G}_i\}$ **FFN parameter**	**Value**
Number of layers	1
Activation	tanh

Table 14: $\{\mathcal{G}_i\}$ FFN architecture

MoE model parameter	$\{\mathcal{G}_i\}$ **FFN hidden size**	**Embedding hidden size**	**Uses BERT encoder**	**Pre-trained BERT**
MoE-1	32	128	No	N/A
MoE-2	128	128	No	N/A
MoE-3	256	768	Yes	Yes
MoE-4	256	768	Yes	No

Table 15: MoE parameters

B.6 Expert Label Functions

During EXP2, we define several types of language experts whose utterance outputs can constitute to smooth bot responses when used in dialogue interactions with users.

The first type of experts we aim to create is *sentiment-based*, because a fluent expression of emotions is important for creating a sense of understanding in human conversations. To quantify the emotional tone of a bot utterance, we use a state-of-the-art (e.g., RoBERTa Sentiment (Liao et al., 2021)) sentiment detector, which outputs 2 sets of prediction probabilities – (i) whether a bot utterance is of positive, neutral, or negative sentiment; (ii) whether the bot utterance falls into any of the 4 more-refined emotions: {joy, optimisim, sadness, anger}.

Neglecting the neutral sentiment output, we define 6 sentiment labeling functions: $\ell_{\text{pos-sent}}(Y)$, $\ell_{\text{neg-sent}}(Y)$, $\ell_{\text{joy}}(Y)$, $\ell_{\text{optimism}}(Y)$, $\ell_{\text{anger}}(Y)$, $\ell_{\text{sadness}}(Y)$, which outputs a score that depends on sentiment prediction probability of any candidate bot utterance. We can also create a *sentiment-empathetic* expert whose utterances will be coherent with the user's sentiment.

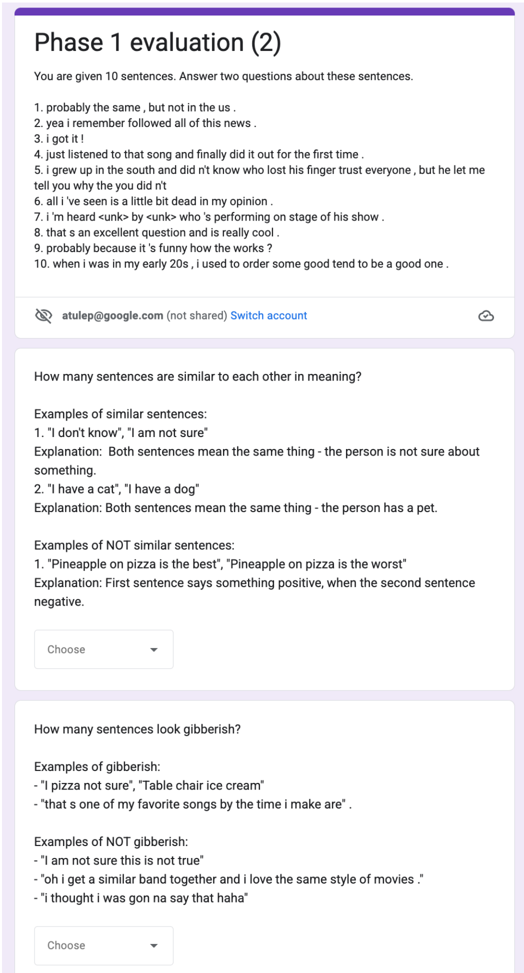
RoBERTa gives scores for each of the 6 categories listed above, which then are summed weighted by the coefficients in Table 16 to produce the *sentiment-based* score:

Sentiment	**Coefficient**
pos-sent	0.5
neg-sent	-0.5
joy	0.5
optimism	1
sadness	-1
anger	-0.5

Table 16: Sentiment Based Label Coefficients

To quantify sentiment coherence between the user and bot, the labeling function $\ell_{\text{sent-coh}}(\mathbf{X}, Y)$ calculates the cosine similarity of user-bot sentiment embeddings (which corresponds to the logit vector of the first set of RoBERTa predictions) in the conversation history. Concretely, it's implemented using tf.keras.losses.cosine_similarity (https://www.tensorflow.org/api_docs/python/tf/keras/losses/cosine_similarity).

The second type of experts of interests is *engagement-based*, whose goal is to encourage user's participation in the conversation. One primary candidate is the *question* expert, because having a bot that is appropriately inquisitive demonstrates the system's attentiveness to users, and thereby also increases user's responsiveness. To characterize whether a bot utterance is linked to a question, we define the label function $\ell_{\text{question}}(Y)$ which outputs 1 when both a question word and a question mark are present and 0 otherwise. Another engagement skill of interests is *exploration*, i.e., the expert is able to change the tone or topic of conversations to avoid having a stale conversation. To measure the amount of exploration, in the label function $\ell_{\text{exp}}(\mathbf{X}, Y)$ we utilize a sentence encoder (e.g., USE (Cer et al., 2018)) to encode both the conversation history and the bot utterance and output the negative cosine similarity between these two embeddings.



Phase 1 evaluation (2)

You are given 10 sentences. Answer two questions about these sentences.

1. probably the same , but not in the us .
2. yea i remember followed all of this news .
3. i got it !
4. just listened to that song and finally did it out for the first time .
5. i grew up in the south and did n't know who lost his finger trust everyone , but he let me tell you why the you did n't
6. all i 've seen is a little bit dead in my opinion .
7. i 'm heard <unk> by <unk> who 's performing on stage of his show .
8. that s an excellent question and is really cool .
9. probably because it 's funny how the works ?
10. when i was in my early 20s , i used to order some good tend to be a good one .

atulep@google.com (not shared) Switch account

How many sentences are similar to each other in meaning?

Examples of similar sentences:
1. "I don't know", "I am not sure"
Explanation: Both sentences mean the same thing - the person is not sure about something.
2. "I have a cat", "I have a dog"
Explanation: Both sentences mean the same thing - the person has a pet.

Examples of NOT similar sentences:
1. "Pineapple on pizza is the best", "Pineapple on pizza is the worst"
Explanation: First sentence says something positive, when the second sentence negative.

Choose

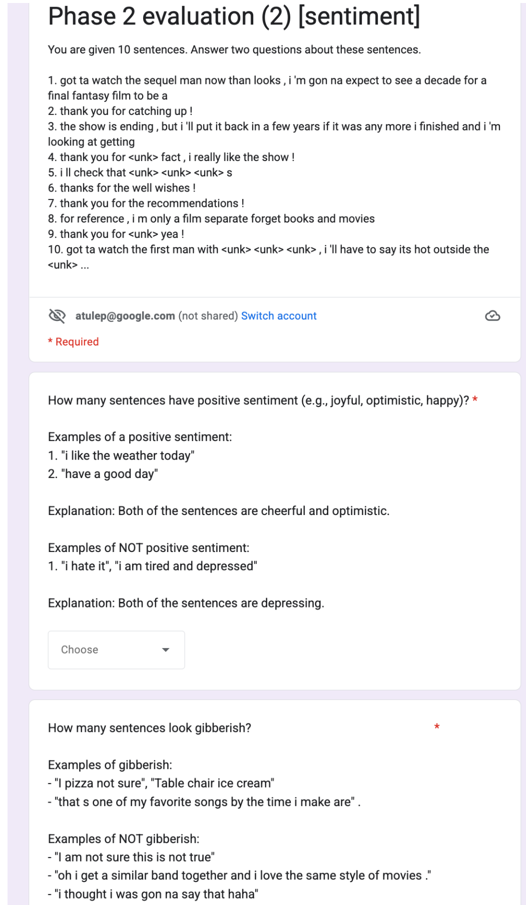
How many sentences look gibberish?

Examples of gibberish:
- "I pizza not sure", "Table chair ice cream"
- "that s one of my favorite songs by the time i make are" .

Examples of NOT gibberish:
- "I am not sure this is not true"
- "oh i get a similar band together and i love the same style of movies ."
- "i thought i was gon na say that haha"

Choose

(a) Phase 1 Raters' Evaluation Template



Phase 2 evaluation (2) [sentiment]

You are given 10 sentences. Answer two questions about these sentences.

1. got ta watch the sequel man now than looks , i 'm gon na expect to see a decade for a final fantasy film to be a
2. thank you for catching up !
3. the show is ending , but i 'll put it back in a few years if it was any more i finished and i 'm looking at getting
4. thank you for <unk> fact , i really like the show !
5. i ll check that <unk> <unk> <unk> s
6. thanks for the well wishes !
7. thank you for the recommendations !
8. for reference , i m only a film separate forget books and movies
9. thank you for <unk> yea !
10. got ta watch the first man with <unk> <unk> <unk> , i 'll have to say its hot outside the <unk> ...

atulep@google.com (not shared) Switch account

* Required

How many sentences have positive sentiment (e.g., joyful, optimistic, happy)? *

Examples of a positive sentiment:
1. "i like the weather today"
2. "have a good day"

Explanation: Both of the sentences are cheerful and optimistic.

Examples of NOT positive sentiment:
1. "i hate it", "i am tired and depressed"

Explanation: Both of the sentences are depressing.

Choose

How many sentences look gibberish? *

Examples of gibberish:
- "I pizza not sure", "Table chair ice cream"
- "that s one of my favorite songs by the time i make are" .

Examples of NOT gibberish:
- "I am not sure this is not true"
- "oh i get a similar band together and i love the same style of movies ."
- "i thought i was gon na say that haha"

(b) Phase 2 Raters' Evaluation Template

Figure 3: Evaluation Template for Running Human Rater Experiment to Study Semantic Diversity and Expert Construction in Language Models.

B.7 RL Rewards

For EXP3, the reward for RL-based dialogue management is defined as $r(\mathbf{X}, a, X_+) = \lambda_1 \ell_{\text{sent}}(X_+) + \lambda_2(\ell_{\text{sent}}(X_+) - \frac{1-\gamma}{1-\gamma^L}\sum_{l=0}^{L-1}\gamma^l \ell_{\text{sent}}(X_l))$, where the linear combination weights $(\lambda_1, \lambda_2) = (0.75, 0.25)$ correlate with Ghandeharioun et al. (2019), and $\ell_{\text{sent}}(X)$ is the same RoBerTa-based sentiment labeler as in EXP2, which assigns a score from $[-1, 1]$ that is proportional to the positive sentiment and inversely proportional to the negative sentiment prediction probabilities. Intuitively, λ_1 assigns weight to *sentiment-based* score of the next user response X_+. λ_2 assigns weight to transition of user sentiment, which we define as the difference between *sentiment-based* score of next user response and *discounted sentiment-based* score of the current conversation.

B.8 Human Evaluation Experiments

We recruited 80 workers to provide a total of 1250 ratings of the bots' quality, in terms of fluency, diversity (phase 1), and the characteristics of the experts (phase 2). Evaluating these language models with humans particularly tests these models' capabilities on generalization, since humans have the final say on judging whether a model response is natural or not. We consider three types of human annotation: fluency, diversity, and relevance w.r.t. expert skills. Annotators are asked to evaluate the fluency, diversity, and expert skills of each individual sample on a scale of 0 to 1. For example, in the fluency rating 0 corresponds to "not fluent at all" and 1 corresponds to "very fluent". We obtain 500 annotations to evaluate fluency/diversity (phase 1) and 750 annotations to compute relevance on expert skills (phase 2).

Primitive Discovery (Phase 1) To evaluate the quality of representation space learned during primitive discovery, we conducted human evaluations on two metrics: (i) diversity and (ii) fluency. Let N be the number of input-output sentence pairs used for evaluating an arbitrary language model, $S(N)$ be the number of similar sentences (each with diversity score below 0.5) out of N total sentences, $G(N)$ be the number of incomprehensible sentences (each with fluency score below 0.5) out of the total of N sentences. Then, diversity metric is given by $(1 - S(N))/N$, and the fluency metric is given by $(1 - G(N))/N$. To test for generalization, for each language model under evaluation we randomly generated $N = 100$ input-output utterances that has not been seen in training, saved each on a Google form shown in Figure 3a and employed a group of 50 raters to obtain $S(N)$ and $G(N)$ for all the language models.

Expert Characteristics (Phase 2) To evaluate the quality of expert characteristics in a language model, we conducted human evaluations on two metrics: (i) expert skill and (ii) fluency. Expert skill is uniquely defined for each chosen expert based on a specific expert label function (see Appendix B.6 for details). In particular, let N be the number of input-output sentence pairs used for evaluating an arbitrary language model, $S_{\text{skill type}}(N)$ be the number of sentences appropriate to a skill type (each with score above 0.5) out of the total of N sentences, then skill metric is given by $S_{\text{skill type}}(N)/N$. Similar to the Phase 1 evaluation, let $G(N)$ be the number of incomprehensible sentences out of the total of N sentences, then the fluency metric is given by $(1 - G(N))/N$. To test for generalization, for each expert skill and each language model under evaluation we randomly generated $N = 50$ input-output utterances that has not been seen in training, saved each on a Google form shown in Figure 3b and employed a group of 30 raters to obtain $S_{\text{skill type}}(N)$ and $G(N)$ for all the language model and skill pairs.