Deep Reinforcement Learning-based Authentic Dialogue Generation To Protect Youth From Cybergrooming

Anonymous ACL submission

Abstract

Cybergrooming is defined as a crime towards potential victims, especially teens, by building close personal relationships with them with the purpose of sexual exploitation via online media. Cyber or online sexual grooming has been recognized as a serious cyber crime. However, there have been insufficient programs to proactively protect the youth from cybergrooming. In this work, we present a generative chatbot framework, called SERI (Stop cybERgroomIng), that can generate simulated conversations between a perpetrator chatbot and a potential victim chatbot. To realize the simulation of authentic conversations in the context of cybergrooming, we take deep reinforcement learning (DRL)-based dialogue generation for authentic simulation of the conversations between a potential victim and a perpetrator (i.e., cybergroomer). The design of the SERI is motivated to ensure a safe and authentic environment to strengthen the youth's precautionary awareness of cybergrooming while any unnecessary ethical issues (e.g., the potential misuse of the SERI) are removed or minimized. We developed the SERI as a preliminary platform that can deploy the perpetrator chatbot to interact with human users (i.e., youth) to observe youth users' responses to strangers or acquaintances and collect the reactions when the youth users are asked for private or sensitive information by the perpetrator. We evaluated the quality of conversations generated by the SERI based on open-source, referenced, unreferenced metrics, and human evaluation.

1 Introduction

As more advanced social media technology has been evolving than ever, people's communication patterns have changed and shown more active interactions with other people in online worlds than offline worlds. The Internet and online social media have brought countless benefits to our life. However, the proliferation of online social media also brought various social problems due to their easy access and deceptive exploitation.

Cybergrooming is one of the well-known online social deception attacks in online social media (Guo et al., 2021). Cybergrooming is a practice on the Internet to establish emotional, intimate, and trusting relationships with potential victims, usually children and teenagers, and use them for online sexual abuse or exploitation, which often leads to offline sexual crimes (Choo, 2009; Marchenko, 2017; Vaswani et al., 2017). CyberTipline (Koons Family Institute on International Law and Policy, 2017) reported that about 60,000 cases were received for sexual purposes on seducing children in cyberspace from 1998 to 2003 in the USA. Due to the unpredictable harmfulness of cybergrooming, some studies have been conducted to investigate the special properties of cybergrooming and develop effective tools to detect predators and the related online sexual abuse behaviors (Anderson et al., 2019; Bours and Kulsrud, 2019; Fauzi and Bours, 2020). However, they mainly focused on detecting predators by analyzing malicious conversations collected from online chatting rooms.

Detecting predators is more reactive than proactive, while it is not highly challenging to detect cybergrooming based on the perspective of normal adults due to their obvious signals. However, due to the unique vulnerability of the youths as teenagers under puberty, a proactive approach is more important to protect our youths from becoming victims by cybergroomers. Due to this reason, taking more proactive approaches to protect the youths from cybergrooming is more critical. In addition, this proactive protection program can contribute to enhancing the sensitivity and awareness of potential victims to the cyber and online grooming situations. Therefore, this work is motivated to develop a generative chatbot framework that can be used for the proactive cybergrooming protection program. This framework is designed to provide

authentic dialogues between the cybergroomer and a potential victim (i.e., a teen) without any ethical issues that might be introduced in education or training programs dealing with sexual abuse or exploitation via online platforms. Our developed generative chatbot framework is named 'SERI' for Stop cybERgroomIng. We developed the SERI as a pre-stage phase before it is deployed to a real human youth user to ensure the provision of a safe and authentic dialogue environment. Therefore, the SERI aimed to provide a safe cybergrooming protection program environment for the youth user to involve an authentic dialogue with a stranger or acquaintance and learn how to respond to such a person asking for sensitive or private information.

While developing the SERI, we faced the following **research challenges**:

- Unlike general conversations between an average adult and a teen, the perpetrator's words are goal-driven and tend to lead a conversation with a potential victim. Ultimately, the perpetrator aims to meet the potential victim in person and exploit the relationship to commit a serious, potential sexual crime. Thus, the perpetrator often takes multiple stages, such as establishing a trust relationship with a potential victim in the initial conversation, gradually escalating its stage to obtaining private information, and ultimately meeting up with the victim in person. However, no prior work has addressed such goal-oriented conversations in the context of cybergrooming.
- A lack of proper datasets has been a non-trivial hurdle in developing a generative chatbot generating authentic conversations. Most related work to online sexual exploitation has used the Perverted Justice (PJ) dataset (Perverted Justice Foundation Inc., 2020a), the only publicly available dataset that the chatbot can mimic. The PJ dataset contains the chatlogs between cybergrooming perpetrators and professionally trained volunteers playing the role of potential youth victims. However, the limited volume of the PJ dataset (i.e., 100 sets of conversations) as well as a lot of noises, such as emojis, slangs, short abbreviations, unsegmented words, or URLs, have been a major challenge to generate high-quality conversations with high logical flows, fluency, and human-like languages.

Under the research challenges above, we made the following **key contributions**:

1. When training the perpetrator chatbot, we employed deep reinforcement learning (DRL) to generate authentic, strategic dialogues where the perpetrator has a clear, ultimate attack goal to achieve offline sexual exploitation. By applying rewards matching a target stage and the corresponding occurrence generation, we augmented the quality of the dialogue model that can generate strategic conversation describing the cybergrooming attack behavior.
2. The SERI was trained via two stages based on the T5 (Text-to-Text Transfer Transformer) model (Raffel et al., 2020): (1) pre-training the perpetrator and victim chatbots on the ConvAI2 (The Second Conversational Intelligence Challenge dataset) (Dinan et al., 2019), which is a causal talk public dataset; and (2) fine-tuning those chatbots on the domain-specific PJ dataset. We preprocessed the PJ dataset by text processing tools to remove the informal slangs, abbreviations, unsegmented words, or emojis from those conversations.
3. Based on the four grooming stages identified from the existing conversations, we modeled the perpetrators by those four grooming stages to achieve the ultimate attack goal, which is meeting up with a victim in person. First, we predicted the grooming stage of each dialogue utterance by training a TextCNN (Kim, 2014). Then, the perpetrator chatbot was trained via the T5 model to provide a response in the target stage. The chatbot was supplied with the dialogues and the reward based on the corresponding grooming stage to guide the dialogue generation.
4. We considered an attack stage-based grooming strategy to manage the dialogue generation based on the perpetrator chatbot. When the perpetrator leads the conversation and collects sufficient resources from the current stage, the perpetrator switches to the next stage by starting a trigger utterance to continue its conversation. If the potential victim shows alertness to this cybergrooming, the victim will terminate this conversation, indicating a failure of the cybergrooming attack.
5. We addressed the performance of the SERI framework based on referenced metrics (i.e., BLEU (Post, 2018), ROUGE (Lin, 2004), and BERTScores (Zhang et al., 2020)) and unreferenced metrics, such as perplexity score and

MaUde score (Sinha et al., 2020). The results demonstrated the high quality of dialogues generated by the SERI. Remarkably, the human evaluation verified the performance by showing that approximately 37% of dialogues produced by the SERI were preferred over the original dialogues in the PJ dataset.

2 Related Work

Cybergrooming detection. Several traditional Machine Learning (ML) algorithms, such as support vector machine (SVM) (Dhouioui and Akaichi, 2016; Gunawan et al., 2018; Anderson et al., 2019; Fauzi and Bours, 2020), k-nearest neighbors (KNN) (Gunawan et al., 2018), Random Forest (Fauzi and Bours, 2020), Decision Tree (Fauzi and Bours, 2020), fuzzy logic (Anderson et al., 2019), Naïve Bayes (Bours and Kulsrud, 2019) and Neural Network (NN) classifiers (Bours and Kulsrud, 2019; Fauzi and Bours, 2020), have been studied to detect cybergrooming from the online forum or social media platforms, by leveraging the lexical features as well as behavioral features. Former studies have developed cybergrooming attack stages (Winters and Jeglic, 2016) among perpetrators and the victims based on the conversational relationship. Perpetrators usually build a relationship and evolve to a closer stage to realize the cybergrooming crime. While most previous studies aimed at detecting and analyzing features of potential perpetrators (Zambrano et al., 2019), there is no research on identifying the characteristics of potential victims in cybergrooming scenarios.

Chatbot application tools. An early-stage chatbot, named Negobot, was developed to detect and analyze potential pedophiles in the social networks (Laorden et al., 2013). A game-theoretic reward can push the chatbot toward the next grooming stage or keep the current stage. In recent years, pre-training language models, such as BERT (Devlin et al., 2019) and GPT (Radford et al., 2018, 2019; Brown et al., 2020), and sequence-to-sequence models, such as BART (Lewis et al., 2019) and T5 (Raffel et al., 2020), have demonstrated their superior capabilities in natural language understanding and generation from the large-scale data training. Several chatbot programs have explored the pre-training language models for conversation generation. For example, the DialoGPT (i.e., dialogue generative pre-trained transformer) (Zhang et al., 2019) fine-tuned GPT-2 on a large-scale conversation dataset to generate coherent and diverse conversations. TransferTransfo (Wolf et al., 2019) also extends GPT-2 with a multi-task objective, combining several unsupervised prediction tasks. However, no previous chatbots have been developed to avoid cybergrooming by simulating conversations between a cybergroomer and a victim.

DRL-based conversation generation. As a typical approach to learn efficient and effective dialogue strategies, Reinforcement Learning (RL), such as Q-learning or SARSA (state-action-reward-state-action), has been commonly used (Levin et al., 1997, 1998; Singh et al., 1999; Roy et al., 2000; Daubigney et al., 2012) to identify optimal dialogue strategies providing high-quality conversation with minimum retrieval cost. Recently, deep reinforcement learning (DRL) has been applied for dialogue generation (Li et al., 2016; Peng et al., 2018) to improve the informativity, coherence, interestingness, and ease of answering. DRL has been used to evaluate emotion (Lan et al., 2021; Zhang et al., 2021), help patients (Rofi'ah et al., 2021), evaluate the interactive RL (IRL) method to offer affordable and faster evaluation, or to generate the dialogue style transfer based on the GPT-2 and BART seq-to-seq models (Lai et al., 2021). However, no prior work has leveraged RL to generate dialogues to model the behaviors and strategies of online social attackers (e.g., cybergroomers) given their attack goals and intents.

3 The Proposed Generative Chatbot Framework: SERI

Figure 1 provides an architectural overview of the proposed SERI framework. The SERI contains two chatbots with the four components as follows: (1) Training a cybergrooming stage classifier to assign a stage to each perpetrator's utterance in the PJ dataset; (2) Pre-training both chatbots for a perpetrator and a potential victim on the large-scale ConvAI2 dataset; (3) Fine-tuning the two chatbots on the preprocessed PJ dataset, and specifically, the perpetrator chatbot is trained with a DRL policy and a reward that measures how likely the generated utterance is from the target grooming stage; and (4) Advancing the perpetrator chatbot to a higher-level stage to continue the dialogues.

Classifying perpetrators' messages per stage. The previous study (Zambrano et al., 2019) de-

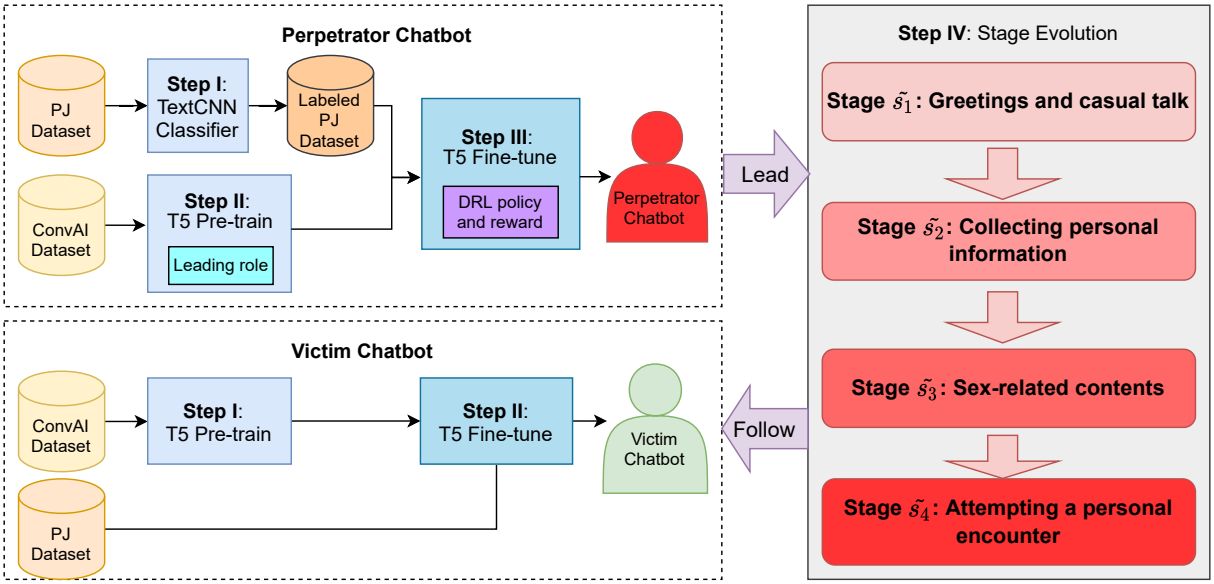

Figure 1: Architecture of the proposed SERI framework.

fined six stages from the perpetrator perspective to indicate the evolution of the cybergrooming conversations, based on a TextCNN (Kim, 2014) trained to predict a stage label for each utterance from the perpetrators. Specifically, given each utterance u, we encode it by the T5 encoder and further feed the contextual representations into a TextCNN model. The output of the convolutional layer after dropout, denoted as $\boldsymbol{u}$, is the contextual representation of u. Then we apply a linear function to classify u to one of the six stages with the softmax function. This stage classifier is optimized by minimizing $\mathcal{L}_C$ based on a categorical cross-entropy loss, which is defined by:

$$\mathcal{L}_C = -\frac{1}{|U|}\sum_{u\in U}\sum_{s\in S} y_{u,s}\cdot\log(\tilde{y}_{u,s}), \quad (1)$$

$$\text{where} \quad \tilde{\boldsymbol{y}}_u = \text{softmax}(\boldsymbol{W}\cdot\boldsymbol{u}+\boldsymbol{b}).$$

Here U is the set of utterances in a dialogue and S is the set of all the target stages. The $\tilde{\boldsymbol{y}}_u$ denotes a vector of probabilities over all stages for u and $\tilde{y}_{u,s}$ is the probability of predicting u with stage s. The $y_{u,s}$ indicates whether s is the true stage label of u by $y_{u,s}=1$ or not by $y_{u,s}=0$. The parameters $\boldsymbol{W}$ and $\boldsymbol{b}$ from the dense layer are learnable.

However, we found the cybergrooming stages proposed in (Zambrano et al., 2019) were not clearly defined because some perpetrators' utterances could fall to multiple stages according to their definitions. Based on our understanding of the grooming stages, we proposed four stages by restructuring the six stages from (Zambrano et al., 2019). To be specific, the new stage $\tilde{s}_1$ is combined from s_1 and s_4, $\tilde{s}_2$ is combined from s_2 and s_3, $\tilde{s}_3$ is the same as s_5, and $\tilde{s}_4$ is the same as s_6. The key conversation contents and topics for the perpetrators covered by each new stage are summarized in Table 1. In the end, through the TextCNN stage classifier and stage consolidation, each utterance in the PJ dataset is assigned a stage label.

Stages	Conversation Content
$\tilde{s}_1$	Greetings, casual talks for initiation of a trust relationship
$\tilde{s}_2$	Private information collection, such as identity as name, age, gender; social relationship as family, school, location; or interests and schedule
$\tilde{s}_3$	Sexual questions or conversations, or sending/requesting sexual pictures/videos
$\tilde{s}_4$	Attempts of in-person contact or requesting online or in-person meeting

Table 1: Cybergrooming stages

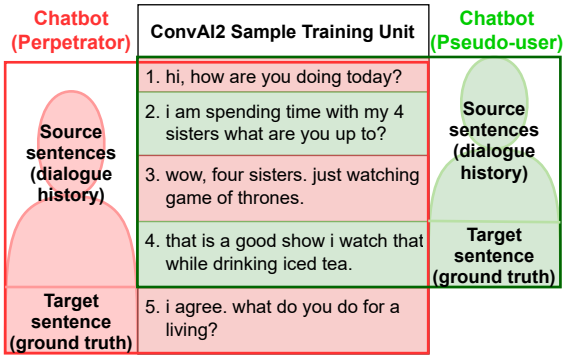

Figure 2: A sample training unit for the perpetrator and pseudo-user (i.e., potential victim) chatbots.

Pre-training the chatbots on the ConvAI2 dataset. We build two chatbots from the T5 model

implemented by PyTorch to play the roles of the perpetrator and potential victim, respectively. Due to the limited size of the in-domain PJ dataset, we first pre-train the T5 model with a large-scale ConvAI2 dataset, which contains broad topic dialogue turns, to improve the fluency of the generated conversations from both chatbots. We noticed that in the cybergrooming conversations, the perpetrators mostly lead the conversations. A similar pattern is also recognized in the ConvAI2 dataset where the conversation is usually between two persons, and the leading person is typically the one who starts the conversation. Thus, to train the perpetrator's chatbot with the ability of leading the conversation, we use the leading person's dialogues in the ConvAI2 to train the perpetrator chatbot and use the other one's responses to train the victim chatbot.

The chatbots training needs to consider the dialogue history from both sides to predict the following utterance response. Then we concatenate four or five consecutive utterances as a training unit[1] and set the last utterance as the prediction target (i.e., ground truth response). The three or four preceding sentences are the training input (i.e., dialogue history) for the victim and perpetrator chatbots, respectively, because we allow the perpetrator's content as the beginning of both chatbots. Figure 2 shows an example of two training units for a perpetrator (red box) and a victim chatbot (green box). In the pre-training phase, given a source dialogue history as x, we have the T5 model (Raffel et al., 2020) to generate a response by optimizing the following objective:

$$\mathcal{L} = -\sum_{i} \log P(y_i|y_{i-k}, \ldots, y_{i-1}; x; \Theta), \quad (2)$$

where Θ is the set of the T5 model parameters and y_i is the i-th token of the response utterance.

Fine-tuning the victim chatbot on the PJ dataset. The fine-tuning of the potential victim chatbot follows the same procedure of the pre-training phase but on the domain-specific PJ dataset. The goal of fine-tuning is to shift the victim chatbot to generate cybergrooming responses.

Fine-tuning the perpetrator chatbot on the PJ dataset with a DRL policy. The perpetrator will take strategies to gradually level up the grooming stages in Table 1 to build a trust relationship with a victim and complete the ultimate grooming goal. As a result, the perpetrator chatbot is able to generate stage-related conversations during the fine-tuning on the PJ dataset. We optimize a DRL policy to generate the utterances closer to the intended stage.

State. A state is denoted by the two previous dialogue turns, and it contains four consecutive utterances $[u_1, u_2, u_3, u_4]$. The dialogue history is further vectorized by feeding the concatenation of u_1 to u_4 into a T5 encoder.

Action. An action is a dialogue utterance to generate. The action space can be considered unlimited since any length sequences within the max-length hyper-parameters can be generated.

Reward. We implement a classification confidence based reward to encourage the chatbot to follow the expected grooming states. We train the stage classifier TextCNN in the previous section and use it to evaluate how well the generated sentence $\mathbf{y}'$ matches the target stage. The confidence of the stage classifier is estimated by:

$$p(s|\mathbf{y}') = \text{softmax}(\text{TextCNN}(\mathbf{y}', \theta)), \quad (3)$$

where $\mathbf{y}'$ represents the generated sentence, $p(s|\mathbf{y}')$ denotes the probability distribution over all the target stage labels, and θ are the parameters of the stage classifier, fixed during fine-tuning. The reward is obtained by:

$$R = [p(s_i|\mathbf{y}')], \quad (4)$$

where $\mathbf{y}'$ is the generated target sentence sampled from the model's distribution at every time step in decoding and s_i is the correct stage from the ground truth.

Gradients and objectives. The reward is used for learning a policy. The policy gradient is given by:

$$\nabla_\Theta J(\Theta) = E[R \cdot \nabla_\Theta \log P(\mathbf{y}^\mathbf{s}|\mathbf{x}; \Theta)], \quad (5)$$

where R is the stage classifier reward, $\mathbf{y}^\mathbf{s}$ is sampled from the distribution of model outputs at every decoding time step, and Θ is the model parameters.

The overall objectives for ϕ are the combination of the loss of the T5 model in Eq. (2) and the policy gradient of the reward in Eq. (5). We test multiple candidate ratios between the two items and identify that 1:0.3 is the best ratio between the loss of the T5 model in Eq. (2) and the policy gradient of the reward regarding our metrics. Figure 3 summarizes the procedures of estimating the loss after integrating the DRL into our model.

[1]For training perpetrator chatbot, we take five consecutive utterances as a training unit, while for training the victim chatbot, we use four.

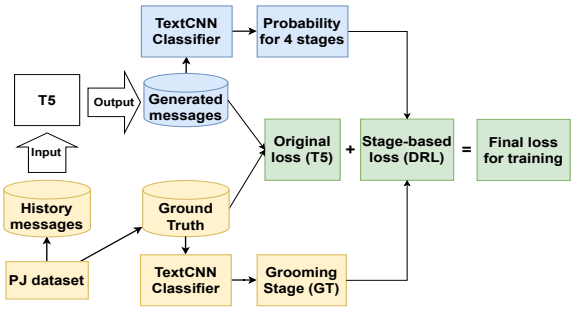

Figure 3: Calculation of loss after integrating DRL.

Stages	Trigger Sentence of Each Stage
$\tilde{s}_1$	hi , how are you doing today ?
$\tilde{s}_2$	you parents know you be chatting with me ?
$\tilde{s}_3$	how many pictures you have , any sexy ?
$\tilde{s}_4$	what will we do if you meet me ?

Table 2: Trigger sentences of the four relationship stages.

Output filtering. After the fine-tuning, to assure the generation of consistent and logically smooth (i.e., human-like) conversations, each chatbot is allowed to produce five candidate utterances every time. We can choose the best utterance based on the *connectivity* scores of the five occurrences to the previous utterance and the *similarity* score to the previous utterance. The *connectivity* scores are computed based on the pre-trained BERT next sentence prediction function and can ensure the consistency of a response utterance with the previous contexts. The *similarity* scores are computed from the Semantic Textual Similarity model (Reimers and Gurevych, 2019) and can maintain the diversity of a generated utterance to prevent duplicate generation. Given a different scale of the connectivity score and the similarity score, we find that 1:3 is the best ratio between them for output filtering.

Stage evolution of the perpetrator chatbot. The perpetrator chatbot not only generates utterances close to a stage but also evolves the grooming stage to a higher level after maintaining a sufficient number of dialogue runs (e.g., 20). For example, if the chatbot stays at stage $\tilde{s}_1$ for 20 rounds, including 10 perpetrator responses and 10 victim responses, this perpetrator will move forward to stage $\tilde{s}_2$. Each stage will start with a trigger sentence (see Table 2), a trigger sentence can direct the conversation into the topic of a specific stage. Each grooming stage evolves based on a trade-off between the risk and benefit to the perpetrator. That is, if the perpetrator is too aggressive, the victim may be aware of the malicious intents and terminate the conversation to cause a failure of the cybergrooming attack. Otherwise, the user can continue the conversation while the perpetrator may not be able to make good grooming progress.

Parameter	Value	Parameter	Value
Learning rate (γ)	$5e^{-5}$	Epochs	4
Epsilon (ε)	$1.0e^{-6}$	Batch size	8
Warmup steps	500	GPU	Yes
Early stopping	0	Vocabulary	T5-base
T5 loss:DRL loss	1:0.3	Diversity coef	3

Table 3: Parameters and their default values used for the SERI framework.

4 Experiment Setup

Datasets. Two chatlog datasets are used in our project. The ConvAI2 dataset (Dinan et al., 2019) is a two-person casual chat dataset in the JSON format with topic labels. We collect 2,000 dialogues with more than 60,000 utterances under the "history" label from the ConvAI2. We manually downloaded each piece of the PJ dataset from the official PJ website (Perverted Justice Foundation Inc., 2020a) in HTML format. It consists of 100 grooming conversations with more than 100,000 chat records between real perpetrators and the professionally trained volunteers acting as potential victims (Perverted Justice Foundation Inc., 2020b). The PJ dataset is split into the training, validation, and testing sets randomly following the ratio of 8:1:1. All other key parameters in our SERI framework are listed in Table 3.

Data cleaning. Due to the well-organized structure, the ConvAI2 dataset is directly suitable for the pre-training steps of our chatbots. However, the PJ dataset is noisy with Emojis, Mentions, URLs, or Hashtags. As data preprocessing, we removed the noises by a regular expression-based Python library 'Preprocessor.' This dataset also has plenty of informal languages, such as lexical slangs, and has long consecutive words omitting space separators. We applied 'wordsegment' library to segment those consecutive words by adding essential spaces in Python. Similarly, the lexical slangs can be normalized with MoNoise (van der Goot, 2019), a state-of-the-art lexical normalization model.

Metrics. The performances of the SERI chatbots are evaluated in terms of the quality of automatic dialogues (Finch and Choi, 2020). We conduct evaluations by referenced metrics, unreferenced met-

Role	BLEU	ROUGE	BERTScore
	Max:100	Max:1	Max:1
Perpetrator	2.556	0.091	0.830
Victim	2.688	0.106	0.827

Table 4: BLEU, ROUGE, and BERTScore-based analysis for the conversations generated by the SERI.

	Perpetrator	Victim
Ground truth dialogues	315.93	477.82
Generated dialogues	124.82	188.97

Table 5: Perplexity score-based analysis.

rics, and human evaluations. The referenced metrics are commonly known as BLEU (Post, 2018), ROUGE (Lin, 2004), and BERTScore (Zhang et al., 2020). BLEU calculates penalty based on the length of the generated sentence and precision of n-gram between generated sentence and references. ROUGE also calculates the recall of n-gram. BERTScore is a metric based on the pre-trained BERT model, computing BERT embeddings and pairwise cosine similarity between generated sentence and reference. We compare the three metrics of the utterances of the proposed SERI against the ground truth utterances in the PJ dataset where the higher measures are better.

The unreferenced metrics are perplexity and MaUde scores (Sinha et al., 2020). The perplexity score is an indicator of how to easily understand a given sentence, where a lower perplexity score represents higher fluency. The MaUde score can judge the language quality in multiple aspects, such as fluency, reasonableness (i.e., logical flow), or repetition avoidance.

Human evaluation is conducted by two graduate students and one NLP expert. Each participating person completes all the evaluations questions. The questions are prepared by randomly selecting 200 conversation snippets with four dialogue histories and two candidate utterance responses. One candidate response is from the ground truth PJ dialogue while SERI generates the other response. Humans can decide which one is more consistent and fluent as the response of the history dialogues.

5 Experimental Results & Analysis

Referenced metrics-based analysis. The results of three referenced metrics BLEU, ROUGE, and BERTScore are shown in Table 4. The scores indicate that the perpetrator chatbot has lower BLEU and ROUGE scores compared to the victim chatbot, reflecting a lower similarity between the SERI's dialogues and the ground truth dialogues. This is because most online chats are informal without strict grammar or fluency rules. The higher BERTScore of the perpetrator can be explained by: (1) BERT failed to learn informative contextual representations from many of the functional and uninformative words, such as *yes*, *haha*, or *why*; and (2) The BERTScore is highly sensitive to certain word pairs which fail to capture any meaningful semantics of very short messages.

	Perpetrator	Victim
Ground truth dialogues	0.844	0.862
Generated dialogues	0.853	0.864

Table 6: MaUde score-based analysis based on PJ evaluation dataset.

	Utterance
Context	1: nutting , you miss me 2: ya 3: you better 4: what if i don't ? , lol , jk 5: i'll get you 6: can't get me through the competition duh , i'm not scared of you
Original response	lol, how much you miss me
Generated response	i'm scared of you right now

Table 7: Inter-agreement sample of human evaluation.

Unreferenced metrics-based analysis. The results of perplexity scores from the ground truth dialogues and the SERI generated ones are shown in Table 5. The ground truth dialogues from the PJ dataset show much higher perplexity scores than the SERI's generated dialogues. Since the perplexity score measures the level of easy understanding, the lower perplexity score from the SERI's dialogues means that the original PJ dialogues have more informal expressions, and grammar or logical errors than the SERI's.

Table 6 shows MaUde scores from the ground truth dialogues and the SERI's dialogues under the PJ dataset. Since MaUde score measures the reasonableness of the dialogues, the SERI's dialogues demonstrate a slightly higher MaUde score than the original PJ dataset. This implies that our SERI chatbots can be significantly resistant to some negative effects caused by the informal languages.

Human evaluation analysis. From the human an-

	With Pre-training	W/O Pre-training
BLEU	2.556	2.505
ROUGE	0.091	0.081
BERTScore	0.830	0.829
Perplexity	124.82	140.33
MaUde	0.853	0.850

Table 8: Impact of pre-training on the ConvAI2 dataset.

	With DRL	W/O DRL
BLEU	2.556	2.472
ROUGE	0.091	0.084
BERTScore	0.830	0.829
Perplexity	124.82	118.93
MaUde	0.853	0.8333

Table 9: Influence of DRL.

notators' responses, at least two annotators agreed 74 SERI produced sample responses out of the total 200. This count achieves a 37% success rate of the Turing test (Turing, 2009), which demonstrates the SERI's promising role in dialogue generation. We provide the SERI response that received unanimously positive response by annotators in Table 7 and show the inter-agreement sample response of human evaluation.

Impact of pre-training and DRL. As shown in Table 8, the dialogue generated by the model with pre-training on the ConvAI2 dataset shows a better performance with all five metrics, compared to the model without pre-training. As shown in Table 9, we observed that overall the model with DRL outperforms the model without DRL. The model with DRL reaches a higher score of BLEU and ROUGE. The aim of using DRL is to simulate the stage strategy of a real perpetrator, which can lead to a higher similarity between the ground truth and generated conversations. Although the model with DRL shows less performance in perplexity score, it was not significant compared to the perplexity performance in the model without DRL. Although the model with DRL showed lower performance in the perplexity score, it outperformed the model without DRL in the MaUde score. This proves that overall using DRL does not introduce a significant quality loss in the generated text while introducing the perpetrator's goal-driven conversation.

Challenges. We found that there are still some challenges with the existing metrics for dialogue evaluation. First, the existing metrics cannot effectively reflect the logical fluency between one utterance and its history utterances. We observed that if conversations are free from grammar errors, the existing metrics simply give high scores without considering logical flows. Second, the existing metrics cannot show the performance of the domain-specific application, such as our chatbot. That is, any existing metrics could not provide meaningful measures to indicate the grooming effect of the perpetrator's utterances on the vulnerability of the victim to cybergrooming.

6 Conclusions & Future Work

We discover several **key findings** from this SERI framework: (1) Pre-training the seq-to-seq dialogue model on a high-quality general conversation corpus first (i.e., ConvAI dataset) and then fine-tuning it on a target corpus (i.e., PJ dataset) enhanced the performance of the proposed SERI compared to training on the target corpus directly; (2) Preparing and segmenting the ConvAI2 data to train the two chatbots with different training data units can match the role of leading conversations by the perpetrator chatbot; (3) Implementing a grooming stage-based deep reinforcement learning method can encourage the chatbot to generate dialogues in accordance with the evolving stages from the perpetrator perspective; and (4) Evaluating the chatbot by the human evaluation demonstrates a promising Turing test rate of 37% to pick the utterances generated by the SERI.

During the implementation and evaluation in Section 5, we also discussed **limitations** from the current development stage of the chatbot, mostly from the domain-specific PJ dataset that has informal styles and poor readability. Although the raw PJ data were cleaned by the automatic text processing tools, it hinders the improvement of the quality of SERI's dialogue generation. Based on the observation of the PJ dataset, the languages used by the perpetrators and the victims have inherently poor quality, which means their uses of informal languages are the key features that distinguish their conversations from other normal conversations.

We have plans of **future research** to: (1) Consider advanced and more intelligent data cleaning methods to deal with social slangs; (2) Investigate how game theory can optimize the current seq-to-seq model to introduce a perpetrator's strategic conversations; and (3) Develop new metrics that can capture the effectiveness of the perpetrator's occurrences on the vulnerability or resilience of the victim to cybergrooming.

Ethical Statement

We aim to develop a chatbot SERI framework to learn to generate simulated conversations between cybergroomers and potential victims, especially children and teenagers. This will be later integrated into a cybergrooming prevention program to improve the sensitivity and awareness of potential victims to online grooming. As there are no proactive programs available to prevent cybergrooming though it has been a serious concern to the society, this work makes a significant contribution to educating and protecting youths from any online sexual grooming. However, everything has two sides, especially the revolutionary Artificial Intelligence technologies. While recognizing the remarkable benefit and contribution of the SERI, we also admit the potential risks and concerns the SERI might introduce. Here we discuss several regulations and strategies to ensure that the SERI will be properly and ethically used by the educators, parents, and children:

- Given the potential concerns, we will not release the programs and models of the SERI to the public. Instead, we will restrict its access to parties (e.g., Education Institutes, research labs, certified parents) for research and education purposes by request.
- A potential ethical concern of the SERI lies in the inappropriate and sexual languages generated by the chatbots. To solve this problem, we will leverage available resources, such as the profane lexicons[2], and design computational approaches to automatically detect the obscene words in real-time and replace them with moderate ones to avoid any potential bad influence to the users, especially children and teenagers.
- We will also design various monitors and strategies to ensure the safety of the SERI and prevent any potential ethical concerns or risks while delivering it as an education program. For example, we will design automatic approaches to keep track of the conversations between the SERI and the users, detect potential grooming activities, and provide alerts whenever a grooming is about to happen. We will also follow the regulations and standards stated in legal systems, such as GDPR[3], and properly use and store the conversational data.

[2] https://www.cs.cmu.edu/~biglou/resources/

[3] https://gdpr-info.eu/

References

P. Anderson, Z. Zuo, L. Yang, and Y. Qu. 2019. An intelligent online grooming detection system using AI technologies. In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6.

P. Bours and H. Kulsrud. 2019. Detection of cyber grooming in online conversation. In *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6.

T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

K. R. Choo. 2009. *Online child grooming: A literature review on the misuse of social networking sites for grooming children for sexual offences*, volume 103. Canberra: Australian Institute of Criminology.

Lucie Daubigney, Matthieu Geist, Senthilkumar Chandramohan, and Olivier Pietquin. 2012. A comprehensive reinforcement learning framework for dialogue management optimization. *IEEE Journal of Selected Topics in Signal Processing*, 6(8):891–902.

J. Devlin, M. Chang, K. Lee, and K. Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

Z. Dhouioui and J. Akaichi. 2016. Privacy protection protocol in social networks based on sexual predators detection. In *Proceedings of the International Conference on Internet of Things and Cloud Computing*, ICC'16, New York, NY, USA. Association for Computing Machinery.

E. Dinan, V. Logacheva, V. Malykh, A. Miller, K. Shuster, J. Urbanek, D. Kiela, A. Szlam, I. Serban, R. Lowe, et al. 2019. The second conversational intelligence challenge (ConvAI2). *arXiv preprint arXiv:1902.00098*.

M. A. Fauzi and P. Bours. 2020. Ensemble method for sexual predators identification in online chats. In *2020 8th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6. IEEE.

S. E. Finch and J. D. Choi. 2020. Towards unified dialogue system evaluation: A comprehensive analysis of current evaluation protocols. *CoRR*, abs/2006.06110.

F. E. Gunawan, L. Ashianti, and N. Sekishita. 2018. A simple classifier for detecting online child grooming conversation. *TELKOMNIKA*, 16(3):1239–1248.

Zhen Guo, Jin-Hee Cho, Ing-Ray Chen, Srijan Sengupta, Michin Hong, and Tanushree Mitra. 2021. Online social deception and its countermeasures: A survey. *IEEE Access*, 9:1770–1806.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Koons Family Institute on International Law and Policy. 2017. *Online Grooming of Children for Sexual Purposes: Model Legislation & Global Review*. International Centre for Missing and Exploited Children.

Huiyuan Lai, Antonio Toral, and Malvina Nissim. 2021. Thank you BART! Rewarding pre-trained models improves formality style transfer. *arXiv preprint arXiv:2105.06947*.

Rushi Lan, Jing Wang, Wenming Huang, Zhenrong Deng, Xiyan Sun, Zhuo Chen, and Xiaonan Luo. 2021. Chinese emotional dialogue response generation via reinforcement learning. *ACM Transactions on Internet Technology (TOIT)*, 21(4):1–17.

C. Laorden, P. Galán-García, I. Santos, B. Sanz, J. M. Hidalgo, and P. G. Bringas. 2013. Negobot: A conversational agent based on game theory for the detection of paedophile behaviour. In *International Joint Conference CISIS'12-ICEUTE 12-SOCO 12 Special Sessions*, pages 261–270. Springer.

E. Levin, R. Pieraccini, and W. Eckert. 1997. Learning dialogue strategies within the markov decision process framework. In *1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings*, pages 72–79.

E. Levin, R. Pieraccini, and W. Eckert. 1998. Using markov decision process for learning dialogue strategies. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, volume 1, pages 201–204 vol.1.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Jiwei Li, Will Monroe, Alan Ritter, Dan Jurafsky, Michel Galley, and Jianfeng Gao. 2016. Deep reinforcement learning for dialogue generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1192–1202, Austin, Texas. Association for Computational Linguistics.

C. Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

S. Marchenko. 2017. Web of darkness: Groomed, manipulated, coerced, and abused in minutes.

Baolin Peng, Xiujun Li, Jianfeng Gao, Jingjing Liu, and Kam-Fai Wong. 2018. Deep Dyna-Q: Integrating planning for task-completion dialogue policy learning. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 2182–2192.

Perverted Justice Foundation Inc. 2020a. Perverted-justice.com archives.

Perverted Justice Foundation Inc. 2020b. Perverted-justice.com info for police.

M. Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Belgium, Brussels. Association for Computational Linguistics.

A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever. 2018. Improving language understanding by generative pre-training.

A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using siamese BERT-networks. *arXiv preprint arXiv:1908.10084*.

Binashir Rofi'ah, Hanif Fakhrurroja, and Carmadi Machbub. 2021. Dialogue management using reinforcement learning. *Telkomnika*, 19(3).

Nicholas Roy, Joelle Pineau, and Sebastian Thrun. 2000. Spoken dialogue management using probabilistic reasoning. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 93–100.

Satinder Singh, Michael Kearns, Diane Litman, and Marilyn Walker. 1999. Reinforcement learning for spoken dialogue systems. *Advances in Neural Information Processing Systems (NIPS)*, 12:956–962.

K. Sinha, P. Parthasarathi, J. Wang, R. Lowe, W. L. Hamilton, and J. Pineau. 2020. Learning an unreferenced metric for online dialogue evaluation. *ACL*.

A. M. Turing. 2009. Computing machinery and intelligence. In *Parsing the Turing Test*, pages 23–65. Springer.

R. van der Goot. 2019. MoNoise: A multi-lingual and easy-to-use lexical normalization tool. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 201–206, Florence, Italy. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.

G. Winters and E. Jeglic. 2016. Stages of sexual grooming: Recognizing potentially predatory behaviors of child molesters. *Deviant Behavior*, pages 1–10.

T. Wolf, V. Sanh, J. Chaumond, and C. Delangue. 2019. TransferTransfo: A transfer learning approach for neural network based conversational agents. *CoRR*, abs/1901.08149.

P. Zambrano, J. Torres, L. Tello-Oquendo, R. Jácome, M. E. Benalcázar, R. Andrade, and W. Fuertes. 2019. Technical mapping of the grooming anatomy using machine learning paradigms: An information security approach. *IEEE Access*, 7:142129–142146.

Rui Zhang, Zhenyu Wang, Mengdan Zheng, Yangyang Zhao, and Zhenhua Huang. 2021. Emotion-sensitive deep dyna-Q learning for task-completion dialogue policy learning. *Neurocomputing*, 459:122–130.

T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi. 2020. BERTScore: Evaluating text generation with BERT. In *International Conference on Learning Representations*.

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