Collaborative Research: SaTC: Medium: Improving Adolescents' Risk-Coping Skills Against Cybergrooming Using Conversational Agents

1 Research Motivation, Goal, and Key Thrust Objectives

Cybergrooming is the process of perpetrators gaining trust and building an emotional bond with young people for sexual exploitation or abuse via online media. According to the National Center for Missing and Exploited Children (2021) [1], during the fall of 2020, the weekly average reported incidents of child pornography were more than 300,000, and over 500 incidents of online enticement of children for sexual acts were reported. Perpetrators often approach adolescents online strategically, pretending to be someone friendly using online information available (e.g., a new incoming student in town or a mutual friend). Perpetrators then increase the level of engagement with a victim gradually and cautiously to lure them towards sexual engagement, both on and offline. They may spend weeks or months building an emotional bond and trust with a potential victim, leading to the adolescent's desensitization to sexually explicit information for sexual exploitation [2, 3]. Other risk factors, such as low self-esteem, loneliness, bullying, and facing family problems, can make some adolescents much more vulnerable to cybergrooming advances than others [4]. Online perpetrators exploit these vulnerabilities, pretending to be one struggling with academic work, family issues, peer pressure, and/or emotional problems. Cybergrooming must be prevented as it is often a precursor of serious crimes, such as child pornography, statutory rape, or sex-trafficking.

The state-of-the-art advanced artificial intelligence (AI) or machine learning (ML) has enabled detection tools that can be deployed in online social media platforms. Such tools can analyze conversations between teens and other users and identify cybergrooming patterns, such as sexting, asking for sending private information, or pictures. Typically, these surveillance tools are designed to alert parents whenever the tools discover suspicious communications with others. However, this approach does not comprehensively prohibit cybergrooming due to several limitations and challenges. Particularly, training such detection models requires a large volume of ecologically valid (realistic) data from teens, which is not readily available. Data collected from minors might include sensitive and possibly illegal artifacts, such as sexually explicit images classified as child pornography. Further, it is highly challenging to distinguish false positives (i.e., treating benign people as perpetrators) from real perpetrators, making these risk detection systems over-flag online interactions to the point of becoming privacy-invasive to teens and cumbersome for parents. Furthermore, perpetrators' awareness on these detection tools can make their strategies become more advanced to avoid the risk of being caught. Lastly, no risk detection system is fool-proof; therefore, proactive education is the only alternative for ensuring that adolescents know how to appropriately respond to potential cybergrooming advances in a way that keeps them safe online. As such, there is a crucial need to diversify our cybersecurity measures of using AI/ML beyond the application of risk detection and surveillance.

The overarching goal of this project is to develop an AI-based experiential intervention program that raises teens' knowledge and awareness about risk factors to cybergrooming as well as increases self-efficacy to protect themselves from cybergrooming and cope with risky situations. To achieve this, we propose to achieve the following objectives via the three key research thrusts:

[Thrust AI] will aim to develop a generative chatbot framework to prevent adolescents from being victims of cybergrooming. We will consider two types of chatbot frameworks: (a) Two chatbots that simulate agents playing the role of a perpetrator and a potential adolescent victim; and (b) a chatbot simulating the perpetrator agent providing conversations when a real adolescent user is deployed to chat with it. The chatbots with two simulating agents will be used to train the perpetrator to minimize potential ethical issues in using languages with adolescent users. Developing the victim chatbot will allow us to initially improve the performance of the perpetrator chatbot without exposing a premature perpetrator chatbot to adolescents. We will develop both chatbot frameworks to achieve knowledge-aware, controllable, goal-oriented (i.e., goals of an attacker and a user), and their strategic conversation generations. After the perpetrator chatbot is trained through the chatbot framework with two agents, it will be used to adolescents through the research activities designed to involve various stakeholders in a safe and age-appropriate manner in Thrust HCI (Human-Computer Interaction), and the experiential learning intervention will be deployed in Thrust HD (Human Development) for validating the effects of the intervention.

[Thrust HCI] will take a participatory approach to build a chatbot-based experiential intervention program to help adolescents identify social cues associated with online cybergrooming risks and develop ef-

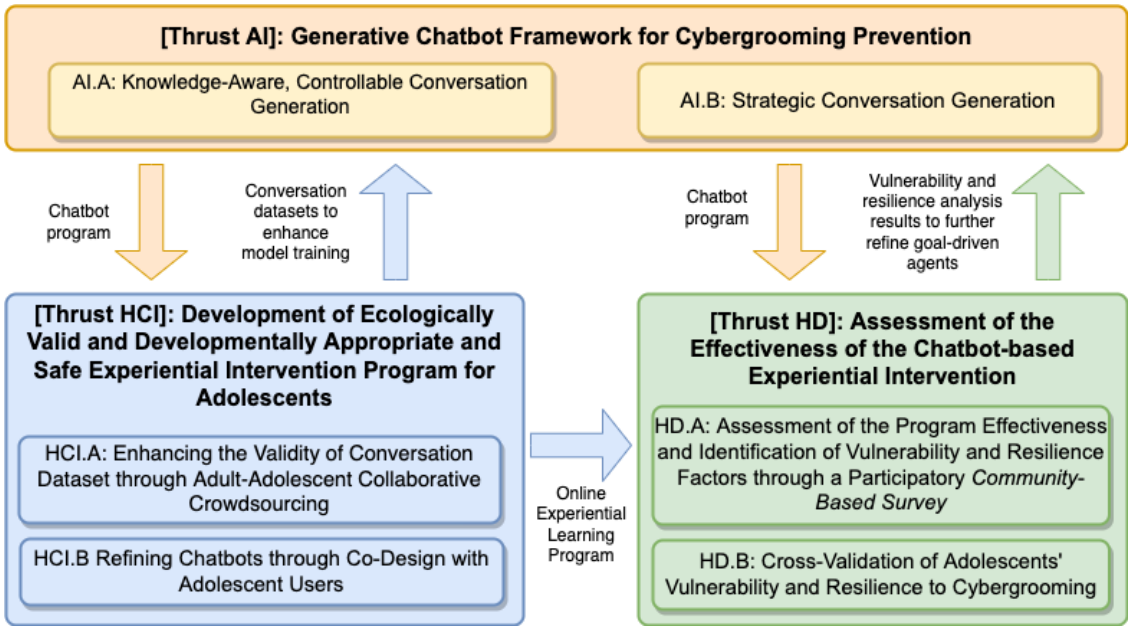

Figure 1: The overview of the key research thrusts and their relationships where AI is Artificial Intelligence, HCI is Human-Computer Interaction, and HD is Human Development.

fective risk-coping strategies for mitigating these risks. We will involve multiple stakeholders, including domain experts, adolescents, and parents, in developing the program and refining the chatbots to be age-appropriate, authentic, and meaningful in simulating the types of cybergrooming interactions teens commonly encounter online. First, we will crowdsource responses from parents and teens to generate a dataset of ecologically valid, safe and unsafe responses to cybergrooming advances that would de-escalate or escalate a cybergrooming situation. Second, we will work directly with teens to evaluate an early prototype of these chatbots and re-design them to be effective and engaging learning tools.

[Thrust HD] will identify the key vulnerability factors of adolescents to cybergrooming and assess the effectiveness of the proposed chatbot-based experiential intervention program via two tasks. First, we will measure the effectiveness of the chatbot-based experiential intervention program on improving their risk-coping skills against cybergrooming-related online hazards. We will use pre- and post-test evaluation design, defined as a before and after assessment to measure whether the expected changes take place in the adolescent participants' understanding and awareness of the victimization processes. Second, we will identify risk, vulnerability, and protective factors that impact the ability of an adolescent participant either to fall a victim to a cyber perpetrator or to recognize and deter such threats. We will also cross-validate the correlations among the key risk, vulnerability, and protective factors based on both traditional survey/interview-based social science methodology and NLP-based, data-driven analyses based on the teens conversation datasets collected from Thrust HCI.

Thrusts HCI and HD will refine the chatbot program developed for experiential intervention by further providing conversation datasets and assessing its effectiveness based on a cross-validation approach. The datasets collected from Thrust HCI will be used in both Thrusts AI and HD to refine the chatbot frameworks and analyze adolescents' risk coping skills against cybergrooming. We summarize the overview of the proposed key research thrusts and the relationships across the key research thrusts in Figure 1.

2 Intellectual Merit & Research Impact

1. The development of an AI-based, proactive cybergrooming experiential intervention program can contribute to protecting adolescents from cybergrooming through the multidisciplinary research efforts. This project will involve five different domains where the five (Co)PIs will contribute their expertise in the areas of AI-based cybersecurity (VT PI Cho), natural language processing (VT Co-PI Huang), human development and sciences (VT Co-PI Kim), human computation and social computing systems (VT Co-PI Lee), and adolescent online safety and human-centered design (UCF PI Wisniewski). The proposed approach

can provide a holistic, integrated view to tackle the severe cybergrooming problem.
2. The proposed research will be the first that investigates the conversations between the perpetrators and victims, discovering the patterns of their acts with a novel temporal act extraction and reasoning framework. In addition, we will develop a new knowledge-aware controllable conversation generation framework that can adaptively select meaningful knowledge and profiles to guide the chatbot generating fluent, coherent, engaging, and human-like conversations.
3. The proposed chatbot framework will be a pioneer work that can generate strategic, goal-driven conversations by using deep reinforcement learning (DRL) and behavioral game theory predicting responses of human users. This is the first that takes a hybrid approach by leveraging both DRL and game theory to develop a proactive, cybergrooming protection program.
4. We are the first to design, develop, and evaluate an AI-based conversational agent as a tool to help adolescents improve their risk-coping skills against cybergrooming-related online risks. Importantly, this work will be done with the utmost care and the input of multiple stakeholders (e.g., teens, parents, and domain experts), given the sensitive nature of the topic. Therefore, a core intellectual merit of this project is the multidisciplinary efforts needed to carry out this work responsibly so that evidence-based practices of adolescents' risk prevention are translated into the design of our system.

3 Relevance to Secure and Trustworthy Cyberspace

Our proposed research is well aligned with the topics of interest designated by the SaTC Core program in terms of the three areas: 'Data Science, Machine Learning (ML), and Artificial Intelligence (AI),' 'Human-Centric Computing,' and 'Social, Behavioral, and Economic Sciences.' Our proposed work requires developing innovative technologies relying on AI, ML, and natural language process (NLP) for the advanced chatbot-based cybergrooming experiential intervention. In addition, we take human-centric computing approaches to help end users and other stakeholders achieve privacy- and security-related goals. We also carefully deal with ethical and societal dimensions of security and privacy, dealing with cybergrooming, one of the serious cybercrimes, with the goal of contributing to building secure and trustworthy cyberspace.

4 Related State-of-The-Art

Knowledge-grounded and controllable conversation generation. A main challenge of conversation generation is the existence of multiple valid responses to a given context. Researchers have imposed various controls to conversation generation, including personality traits [5, 6, 7, 8, 9, 10], sentiment and emotion [11, 12, 13, 14, 15], goals or situations of speakers [16, 17], categories of reacts [18], world facts from Wikipedia and DBpedia [19, 20, 21, 22, 23, 24], and commonsense [25, 26, 27, 28, 29]. However, cybergrooming conversation generation is more challenging as there is no knowledge resource available. In addition, the perpetrator usually leads the conversation, which is different from general chatbot conversation where the chatbot takes a passive role in most cases, such as answering a simple question.

Strategic conversation generation. Reinforcement Learning has been commonly used [30, 31, 32, 33, 34, 35, 36, 37] to identify optimal dialogue strategies providing high-quality conversation with minimum retrieval cost. Deep reinforcement learning (DRL) has been considered for dialogue generation [38, 39] considering informativity, coherence, and ease of answering in the reward functions and evaluated their model on diversity, length, and human judges. Game theory also has been used to generate conversations for maximizing interpretation [40] or finding optimal utterances for games [41, 42, 43, 44, 45], or minimizing vagueness [46]. However, no prior work has leveraged DRL- or game-based dialogue generation to model the behaviors of goal-driven online sexual perpetrators.

Sociotechnical Intervention for cybergrooming. Cybergrooming has been studied mainly based on algorithmic detection of perpetrators' messages [47, 48, 49, 50, 51, 52, 53, 54, 55]. In practice, however, such algorithms can be only deployed at the price of sacrificing adolescents' privacy as the monitoring tools would need access personal data and share it with their parents [56]. Co-PI Wisniewski has studied adolescent online safety [57, 58, 59, 60, 61, 62] to empower adolescents by supporting self-regulation, communication with parents, and effective education to engage safely with others online. However, the literature has limited understanding on how to take advantage of modern AI/ML to design, develop, and deploy sociotechnical interventions that can empower adolescents beyond surveillance.

Resistance and receptivity to cybergrooming for experiential intervention. A fair amount of literature has discussed the risk factors shared among the victims of cybergrooming [63, 64, 65, 66, 67] as

well as the coping strategies demonstrated by both the victims and non-victims of cybergrooming [68]. A set of common risk factors are being a girl, having low self-esteem, experiencing child sexual abuse, and being willing to meet a stranger in person. In contrast, parental supervision and instructive mediation help adolescents avoid disclosing too much private information online, which can significantly reduce the risk for cybergrooming. Unlike the works above, we will investigate hybrid approaches taking both survey/interview-based and data-driven approaches to identify adolescents' social, psychological, and environmental factors impacting their responses to cybergrooming situations to ensure high validity of our research findings.

5 PIs' Prior Research Related to this Project

NLP-based Cybersecurity Interventions. This project leverages the teams' expertise on the subject matter and the proposed method and is based on the knowledge gap identified by the PIs. PI Cho has studied DRL [69, 70, 71, 72] and game theory [73, 74] to solve various cybersecurity problems. Co-PI Huang has studied knowledge acquisition [75, 76, 77, 78, 79, 80, 81] and natural language generation (NLG) [82, 83, 84, 85, 86, 87]. Cho and Huang have developed a chatbot framework [88, 89] to develop a chatbot-based cybergrooming prevention software, mainly focusing on only improving the fluency of the conversation. Cho, Huang, and Kim, also investigated potential victims' vulnerability to cybergrooming in [90] using NLP tools. We will further refine the chatbot based on Thrust AI and examine the key vulnerability and resilience factors of cybergrooming based on the proposed cross-validation in Thrust HD.

Involving Adolescents and Participatory Approach. The project is also motivated from PI Wisniewski's previous findings that teens are often able to regulate themselves and even benefit from experiencing some level of online risk by learning from their mistakes and developing important interpersonal skills [91, 60]. Co-PI Lee has worked on designing and implementing online social computing systems, some of which involve non-expert users in the field [92, 93, 94, 95]. Lee's technical interventions focus on how technology can be used to facilitate understanding and empathy through collaboration and participation among people from groups that have distinctive goals and motivations [92, 96, 97, 93, 98, 99, 100, 101, 102, 103]. In addition, his experience in facilitating collaboration among crowd workers will be used to propose a new model of involving adolescents in human computation tasks [93, 98, 104, 105, 106, 107].

6 Research Description

6.1 [Thrust AI] Generative Chatbot Framework for Cybergrooming Prevention

Objective: Thrust AI (Artificial Intelligence) will aim to develop a generative chatbot framework that can be used for the cybergrooming prevention program for adolescents. To achieve this, we propose the following two tasks. First, we will develop knowledge-aware, controllable conversation generation to consider their diverse profiles and knowledge at multiple levels of details. Second, we will realize strategic conversation generation by leveraging DRL to consider goal-driven, strategic simulating agents while predicting human users' responses to refine the chatbot framework to provide more intelligent, human-like responses.

Thrust AI: Research Questions

1. *Given the pre-specified profile of each role, how can the chatbot be enforced to generate informative, logically consistent, and "human-like" conversations by leveraging the profiles of each role and various knowledge resources?*
2. *How can vulnerability- and resilience-aware reward in deep reinforcement learning help simulating perpetrator and user agents achieve their respective objectives respectively in the conversation?*
3. *How does the accurate prediction of a human user's action (i.e., utterance) affect the perpetrator agent's performance in maximizing the chances to achieve its objective?*

This thrust will develop two types of the chatbot environments: (1) Simulation Environment I (SE-I): Two simulating agents playing the roles of a perpetrator and a potential victim; and (2) Simulation Environment II (SE-II): Simulating a perpetrator agent to converse with a real teen user. SE-I with two simulating agents is to ensure the perpetrator chatbot's fluency, coherence, engaging, goal-driven, and strategic conversations before the perpetrator chatbot is deployed to a real teen user. To minimize any potential ethical issues that might be caused by its distasteful languages, we will train the perpetrator chatbot to minimize the use of profane languages while delivering ecologically valid, safe cybergrooming scenarios.

6.1.1 Task AI.A: Knowledge-Aware, Controllable Conversation Generation

We will first develop a chatbot framework to simulate the conversations between a perpetrator and a potential victim based on their various profiles and knowledge at multiple aspects, such as acts from cybergrooming scenarios or world fact knowledge. Despite large amounts of resources available for general domains (e.g., Wikipedia and ConceptNet), there is no readily available knowledge resource for cybergrooming. We will automatically acquire and curate grooming-related act knowledge from available corpora. Then, we will integrate the knowledge with various perpetrator and victim profiles to mimic each role's mind in conversations and guide the development of the chatbots, reflecting the context of cybergrooming.

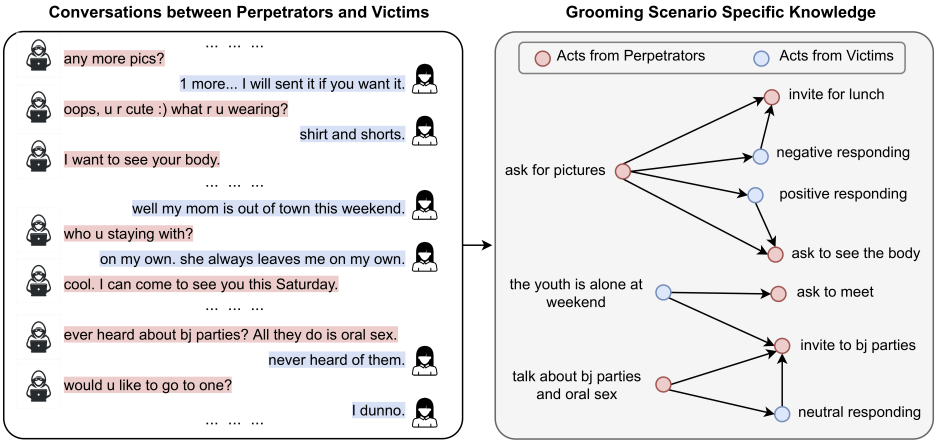

Figure 2: Example Act Knowledge Extracted from Cybergrooming Conversations

Temporal Act Knowledge Acquisition from Cybergrooming Conversations. Conversations between the perpetrators and victims usually involve various grooming acts, such as asking to send nudes or inviting to sex parties. Experienced perpetrators are often skilled at easily establishing trust relationships with adolescents, collecting information about their location and parents' schedule, and incrementally leading to talking about sexual acts or meeting up in person. Such perpetrators' acts usually follow inherent temporal patterns (e.g., asking to see the body usually follows asking for pictures) and are conditional on the responses of the teens they talk with, such as positive, negative or neutral responses as well as other situations from the teens. As shown in Figure 2 using the PJ dataset, if the teens respond positively to the request of sending pictures, the perpetrators tend to ask for sending more sensitive information. Otherwise, shifting to other topics, the perpetrators may ask if the teens are alone at home or ask to meet on weekend. The temporal act knowledge can significantly benefit the chatbots on simulating the minds of perpetrators and developing the grooming process between the perpetrators and victims. However, the temporal act knowledge is usually implicitly embedded in the unstructured conversations without any structured knowledge resources available.

Table 1: CATEGORIES OF CYBERGROOMING ACTS

Label	Category Definition
A_1	Establish a link with the victim when talking about their personal interests.
A_2	Acquire specific information of the victim, related to his/her friends, family, school and social life.
A_3	Show compassion and understanding to gain the victim's trust.
A_4	Inquire about the location, the parents' schedule and the victim's supervision.
A_5	Express feelings of love, care and confidence.
A_6	Adulate the victim to maintain and increase the level of trust.
A_7	Desensitize the victim in the sexual theme, using biological terms.
A_8	Detail the sexual acts that the perpetrator wants to perform with the victim or past sexual experiences.
A_9	Plan a personal encounter with the victim.

We will design a bottom-up approach to automatically acquire and curate the acts of perpetrators and victims from conversation datasets for cybergrooming, such as the Perverted Justice (PJ) dataset [108] which consists of conversations between perpetrators and undercover polices mimicking adolescent victims, and then reason about their temporal relations. Inspired by the communicative intentions of the perpetrators discussed in [109], we define nine categories of acts performed by perpetrators in cybergrooming scenarios, as shown in Table 1. For victims, we define two categories of acts: (1) responding act, such as positive, negative, or neutral responses; and (2) describing personal situation, such as being alone on weekend. Due to a lack of datasets on cybergrooming conversations, it is highly challenging to obtain a large amount of manual annotations to identify the acts of perpetrators and victims. As such, we propose a weekly supervised entailment analysis approach where each candidate act from perpetrators or victims is viewed

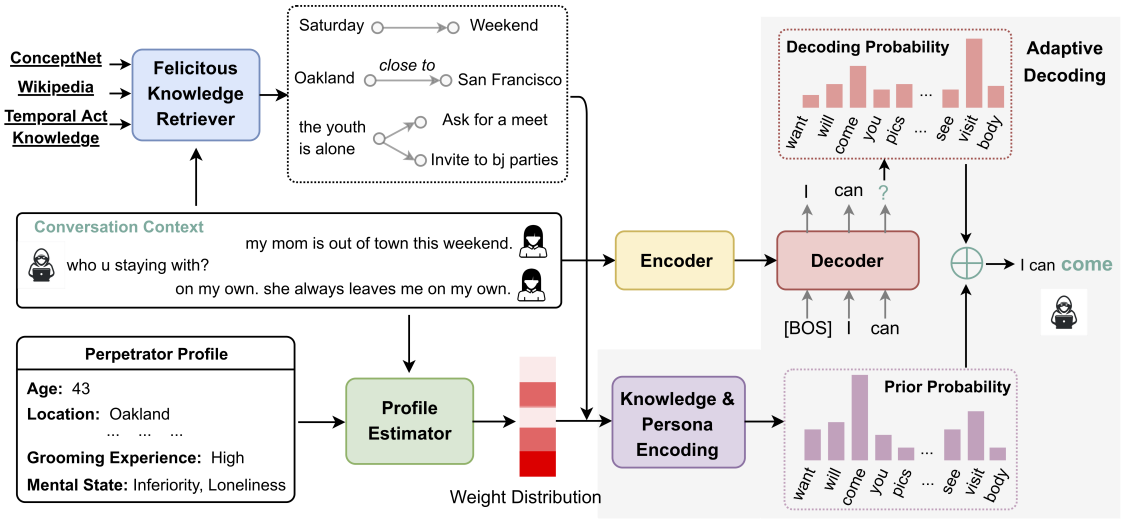

Figure 3: The overview of Knowledge and Persona Guided Conversation Generation Framework

as a premise and we automatically determine the semantic correlation of the premise with the definition of each category. To infer the temporal relations among the discrete acts of perpetrators and victims, we will use a Temporal Graph Transformer (TGT) framework we recently proposed [87], where a new graph neural networks (GNNs) is designed to automatically learn a rich semantic representation for each act and predict the temporal relation (before, after, or vague) of each pair of acts based on their contextual features.

Knowledge and Persona Guided Conversation Generation. When humans converse in reality, they frequently associate the conversation context with various background knowledge in their minds, such as world facts, common sense, or past experiences.

Table 2: PROFILE ATTRIBUTES OF A PERPETRATOR AND VICTIM

Role	Profile Attributes
Perpetrator	Age, gender, location, interest, employment, marital status, mental state, grooming experience
Victim	Age, gender, location, interest, mental state, resilience level.

In addition, conversational strategies and languages used by the perpetrator and victim heavily depend on their psychological and demographic profiles [109]. To realistically simulate the conversations between perpetrator and victim chatbots, we will explore a novel *knowledge and persona-guided conversation generation framework*. We will leverage multi-granularity knowledge resources, including Wikipedia that includes all world facts, ConceptNet which is a human commonsense knowledge base, and the temporal act knowledge acquired from in-domain grooming conversations. As for human profiles, we define a set of persona attributes for both perpetrator and victim, respectively, as shown in Table 2. The value of each attribute, such as grooming experience of the perpetrator, could be a short text (e.g., high, medium, or low) or unknown.

Figure 3 depicts the overview of our chatbot framework, consisting of three innovative components to integrate the background knowledge and persona dynamically to guide the chatbot to generate each utterance. First, given a particular conversation context, the *Felicitous Knowledge Retriever* is designed to identify the most relevant and appropriate background knowledge from various sources. It will be built on the state-of-the-art neural text encoder (e.g., BERT [110]) to encode the context, and GNNs (e.g., graph transformer) to encode each triplet from the multiple background knowledge resources, and select the top-k ranked triplets based on their similarity. At each turn, the various perpetrator or victim profiles may play different roles to the target utterance. For example, as Figure 3 shows, knowing that the victim will be alone this weekend, the perpetrator will propose to meet on Saturday, which is mostly driven by the 'location' (e.g., Oakland is close to San Francisco), 'grooming experience' as well as the 'mental state' of the perpetrator. We further design a *Profile Estimator* to predict a weight for each attribute based on the conversation context. Previous knowledge-guided generation approaches mainly rely on the encoder to integrate the knowledge into the context representation and generate each word based on the probability distribution predicted by the decoder. However, by simply considering the knowledge at the encoding stage, these methods are not efficient in leading the decoder to generate the desired response. We propose a novel *Adaptive Decoding* strategy to first produce a prior probability distribution over all possible words based on

the relevant background knowledge and the weighted combination of all profiles. At each decoding step, we will combine the prior probability with the probability predicted by the decoder, to ensure the chatbot model to fully consider the background knowledge and persona and produce logically consistent utterances.

6.1.2 Task AI.B: Strategic Conversation Generation

GPT [111, 112, 113, 114, 115, 116], BERT [117], or Seq2Seq [118] based approaches have been substantially explored in the chatbot research. However, since the works above mainly aimed to improve the quality of languages in naturalness and fluency, using them cannot meet the purpose of the proposed chatbot-based experiential intervention program to effectively prevent cybergrooming. Game theory has been used to model a pedophile's conversation exhibiting the intent of online sexual exploitation without considering the moves of the opponent (i.e., a potential victim) [119]. This task will develop a strategic conversation framework that helps each agent choose occurrences to achieve its own goal by effectively balancing between possible conflicting goals.

Table 3: CYBERGROOMING STAGES

Stages	Conversation Content
$\tilde{s}_1$	Greetings and casual talks to establish a trust relationship
$\tilde{s}_2$	Collecting private information, such as name, age, gender, location, interests, family, school, or schedule
$\tilde{s}_3$	Asking sexual questions or requests, talking about sexual conversations, or sending sexual pictures/videos
$\tilde{s}_4$	Attempting a personal contact or asking meeting in person

To be specific, the perpetrator agent will want to generate effective but aggressive utterances to meet up a potential victim in person while it may fail its goal if the aggressive utterances make the potential victim be scared and leave. Similarly, the potential victims want to have friendly conversations with new friends while protecting themselves from a potential perpetrator. These conflicting goals will be addressed in the formulation of a reward function in the DRL-based dialogue generation.

Deep Reinforcement Learning (DRL)-based Conversation Generation. Unlike existing RL-based dialogue generation techniques [30, 31, 32, 34, 120, 35, 38, 33, 121, 39, 37] to improve fluency, diversity, and/or conversation length, we will develop DRL-based dialogue generation where both a perpetrator chatbot and a user chatbot (i.e., a potential victim) aim to achieve their respective goal in the conversation, as discussed above. The key components of DRL-based conversation generation process will follow: (1) ***Action***: A player chatbot will generate an action a which is a dialogue utterance (u). The perpetrator will use the so-called *cybergrooming kill chain* describing the process of cybergrooming until the perpetrator achieves the attack goal. We propose the four stages of cybergrooming process by the perpetrator as shown in Table 3. Although the perpetrator will follow the process of the four stages to achieve its goal, the perpetrator will mainly select actions (i.e., u_P) from the current stage $\tilde{s}_i$ for $i = 1$ to 4 while choosing utterances in other stages to maintain fluency. In SE-I, we will simulate a user agent (i.e., a potential victim) based on the resilience level against cybergrooming to increase the adaptive capability of a perpetrator agent to handle a different type of users. The simulated user's actions will be a set of utterances belonging to the user's resilience stage where we consider the three levels of resilience: low, medium, and high (i.e., u_V^L, u_V^M, u_V^H). This will allow the perpetrator agent in SE-II to adaptively respond to a human user depending on the user's resilience level. (2) ***State***: When a dialogue is represented by an alternating sequence of perpetrator and user agents, $p_1, q_1, p_2, q_2, \ldots, p_i, q_i$, where (p_i, q_i) refers to a *turn*, a *state* indicates the two

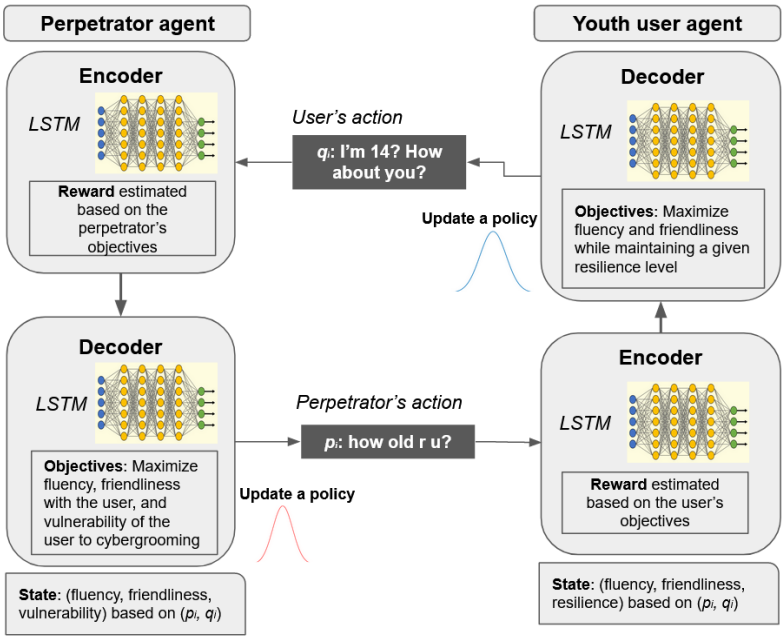

Figure 4: The Core Structure of the Proposed DRL-based Simulating Agents for the Perpetrator and the Teen User.

previous dialogue turns, (p_i, q_i). We will consider the dialogue history to be transformed to a vector with the concatenation of p_i and q_i which is fed into an LSTM (Long Short-Term Memory) encoder model [115, 38]. (3) ***Policy***: A policy refers to the probability distribution of actions given states, $\pi(p_{i+1} \mid p_i, q_i)$, as the form of an LSTM encoder-decoder, defined by its parameters. Simulating agents will make stochastic choices of actions based on the policy. (4) ***Reward***: We denote the simulating perpetrator's and user's reward by r_P and r_U, respectively, indicating the reward by taking a given action (i.e., utterance, u_P and u_U). Although both simulating agents will have their objectives to generate dialogue with high fluency and diversity, they also need to generate goal-oriented dialogue to achieve their objective. The perpetrator's action will be rewarded when the user's response shows higher vulnerability (e.g., sharing more private information) to the cybergrooming while the user keeps engaging the conversation. To this end, we will develop a vulnerability metric to measure the user's vulnerability to cybergrooming as a function of the user's action to the perpetrator's action. We will consider datasets obtained from Thrust HCI and use them to measure adolescents' vulnerability factors to cybergrooming in Thrust HD. To model each agent's multiple objectives to be effectively optimized, we will consider various multi-objective optimization techniques [122].

Based on the correlations identified between resilience/vulnerability and being a victim of cybergrooming in Thrust HD, we will develop a success metric of the user's performance (e.g., resilience). In SE-I, we will use the Text-to-Text Transfer Transformer (T5) [123] and train the T5 using both public conversation datasets (e.g., ConvAI2 dataset [124]) and the PJ dataset [108]. In SE-II, the utterances by the human user may be new to the perpetrator chatbot. We will consider unsupervised dialogue generation where the human user's utterances should be evaluated based on the vulnerability metric, which can guide the perpetrator chatbot to select its best actions to maximize its reward. We will consider both policy-based [125, 126, 127] and value-based [128] DRL algorithms to optimize the learning capability of the chatbots.

Behavioral Game Theoretic Prediction of Human User Behaviors. In a multiagent game, each agent is assumed a rational entity to maximize its utility and accordingly adopt Nash Equilibrium (NE) strategies [129, 130]. However, human players may not choose their NE strategy in the initial play of normal-form games. Predicting a human player's behavior has been significantly studied in *behavioral game theory*, considering humans' cognitive bias and limitations [131]. In SE-II, we will leverage the behavioral game theory to predict the human user's action. For the perpetrator chatbot to choose its optimal action maximizing its utility for the success of cybergrooming, it is critical to accurately predict the human user's action, which will be considered in the perpetrator's utility function. We will leverage machine/deep learning to predict the human user's action in the degree of presenting vulnerability to cybergrooming in the response utterance. Since the occurrence depends on the context of a dialogue, we will identify a set of utterances associated with a predicted vulnerability to each action by the perpetrator chatbot. Since the PJ dataset has the victims' conversation with professionally trained volunteers playing victims, they may intend to lure perpetrators and entrap them, rather than exhibiting natural, perhaps cautious, responses of teens to strangers. We will mitigate this effect by training our models using the conversation datasets collected from Thrust HCI.

6.2 Thrust HCI: Development of Ecologically Valid and Developmentally Appropriate and Safe Experiential Intervention Program for Adolescents

Objective: Thrust HCI will aim to develop an ecologically valid, developmentally adequate, and safe experiential intervention program for adolescents via two tasks. First, we will create a closed-loop through which various stakeholders can iteratively improve the chatbot performance in Thrust AI. Second, we will develop an experiential intervention program that leverages the perpetrator chatbot in Thrust AI to help adolescents enhance their risk-coping skills and build protective strategies against cybergrooming.

Thrust HCI: Research Questions

1. *What are ecologically valid responses to a perpetrator chatbot that would either increase or decrease a teen's vulnerability to cybergrooming advances?*
2. *What are the key design-based and ethical challenges of deploying a perpetrator chatbot for letting adolescents learn how to combat cybergrooming advances?*
3. *How can we overcome these challenges in a human-centered, developmentally appropriate, and provide a meaningful way that empowers adolescents as end users?*

We propose the following two main tasks to address the research questions: First, we will involve adolescents safely in annotation and generation tasks through parent-adolescent collaborative crowdsourcing. Second, we will involve adolescents in participatory designs of an online cybergrooming experiential intervention program for them to cope with risks exposed by the program.

6.2.1 Task HCI.A: Enhancing the Validity of Conversations through Collaborative Crowdsourcing

Addressing the Limitation of the PJ Dataset. The user experience of the chatbot depends on the performance of the conversational agents' model, which itself depends on the training dataset. The PJ dataset [108] used in Thrust AI provides ecologically valid annotated data from real child perpetrators. However, the dataset is limited in multiple ways as adolescents were adults posing as young victims. It is a critical flaw that the PJ dataset only provides victim-based responses designed to entrap a perpetrator and lacks protective strategies as a cybergrooming stage advances. Therefore, one critical gap that is necessary to close is involving key stakeholders in dataset generation and refinement for enhancing the validity of the dataset to enhance the chatbots in Thrust AI.

Ethically Enhancing the Validity of Conversational Agents' Performance. One common approach to enhance the validity of conversational agents' performance is to leverage crowdsourcing [132, 133, 134, 135, 136]. For example, crowd workers can vote among multiple responses suggested by agents or propose a response to a chatbot conversation [136]. However, participation in crowdsourcing platforms is limited to those who are 18 or older, making it difficult to involve adolescents in refining and generating the dataset [137]. In addition, involving adolescents in dataset refinement and generation should be approached carefully due to several considerations. First, adolescents may not know what an appropriate risk-coping response to a cybergrooming advance should be. Second, the responses from teens may contain sensitive content, such as personal information or risky content of a sexual nature. Therefore, adolescents' involvement in dataset refinement is needed to protect participating adolescents. We will create a crowdsourcing task model where adult workers (i.e., parents) can partner and consult with their children to provide ecologically valid responses that would either deescalate or escalate a cybergrooming scenario. The proposed model of adult-adolescent collaborative crowdsourcing is a novel, methodological contribution to the literature of human computation for facilitating local collaboration between a crowd worker and their family members. We will leverage existing collaborative approaches [138, 98] based on adult-adolescent pair crowdsourcing for data collection from adolescents. The benefit of this approach includes diverse responses from adolescents in an ethical and responsible way. Lastly, this process will also benefit the crowd worker (i.e., a parent), giving an opportunity to have discourse on sensitive topics through guided collaborative tasks. We will deploy adult-adolescent collaborative tasks to be only available through the adult crowd workers' consent and their willingness to involve adolescents as they complete the tasks. This procedure will ensure that adolescents' involvement is not only ethical and safe but also generative and empowering. For instance, a message containing a direct sexual solicitation generated by the perpetrator chatbot would not be shared without the parent crowd worker seeing the task prior to their teen and giving consent to the task. The adult crowd worker will take the primary role to decide if they will involve their children or not, which will be an opportunity for us to collect the reasons why they would or would not want to involve adolescents in furthering the data collection tasks (e.g., educational purposes vs. explicit content, risk behaviors). We will explicitly contextualize how data will contribute to developing an experiential intervention to cybergrooming for adolescent workers to understand how their participation will help other adolescents.

We will collect in total four different types of data from the crowdsourcing task: (1) Subjective profile information (Table 2 and its variance in adults and adolescents' responses); (2) generating responses for completing chatbots conversations in various scenarios given the objectives of a perpetrator and an adolescent chatbot; (3) votes on chatbots generated messages for ecological validity evaluation and cybergrooming framework annotation; and (4) responses that parents and adolescents choose to make for perpetrators' risky requests (e.g., sharing private information, request photos, in-person solicitation) for understanding adolescents' vulnerability and risk-coping strategies. The collected data will directly improve the fluency of the chatbots' conversation and generate novel datasets that can be published. For example, if a crowd worker is asked to modify a chatbot's generated utterance to be more ecologically valid, they will be asked to involve their child to modify the message to reflect the linguistic jargon used by teens. The research outcome will be adolescent annotated and generated datasets that other researchers can use.

We will follow the privacy-preserving data collection and storage policy described in the submitted data management plan as well as the protocols described in Section 8. The datasets collected in Thrust HCI will be used to refine the chatbot to be developed in Thrust AI, which will also conduct its performance analysis. The ways that those two stakeholders respond to the chatbots and annotate the adolescent chatbots' utterances can differ. We will explicitly ask parents to employ protective, risk-coping strategies as their main goal is to protect their children from cybergrooming. Adolescent crowd workers may choose to respond in ways they see fit, and other crowd workers will evaluate their responses to assess whether they deescalate vulnerability to cybergrooming or escalate it. The disparity between adolescents' and parents' responses can be used to develop the chatbot framework to reflect the game-theoretic nature of adolescents' online behaviors when conversing with strangers, such as wanting to keep engaged in conversation vs. disengaging to protect themselves as well as how perpetrators chatbots should optimize their strategies given its utility function. The chatbot framework in Thrust AI will be improved by using the dataset provided by parents and adolescents in Task HCI.A.

Although involving adolescents in crowdsourcing is an innovative approach, we will design this phase of the research with extreme care. Co-PI Lee has extensive expertise in designing collaborative tasks in crowdsourcing [139, 98, 93], while Co-PI Wisniewski has deep expertise in involving adolescents in risk-based research [104, 140, 141, 142]. For example, Wisniewski's recent work has surmounted the technical, legal, and ethical challenges of having adolescents users safely and privately donate their Instagram data and annotate their private messages for situations that make them feel uncomfortable or unsafe [143]. Similarly, we will have our crowdsourcing tasks reviewed by the Internal Review Boards of both universities and seek feedback from the domain experts who provided letters of collaboration for this proposal.

6.2.2 Task HCI.B: Refining Chatbots through Co-Design with Adolescent Users

Experiential Intervention Program. In Task HCI.B, we will design and develop an experiential intervention program that combines evidence-based risk prevention strategies for cybergrooming within a web-based version of the chatbot to make it user-friendly and accessible for wider audiences. Scenario-based sexual risk prevention and decision-making programs have shown promise in offline contexts [144, 145]. Therefore, our intent is to extend these proven frameworks to the online context of cybergrooming. PI Wisniewski will work with Dr. Erica Fissel and Jan Edwards, domain experts in cybervictimization and sex trafficking prevention (see their letters of collaboration), to create a web-based training module that will give parents and teens actionable advice on how to respond to potential cybergrooming scenarios in a way that deescalates the situation. PI Wisniewski will also leverage her prior work on privacy and adolescent risk-coping to online risks [60, 58, 146] to inform the design of the training material.

Participatory Design Workshops with Teens. As part of her NSF CAREER Award, Wisniewski leads a Participatory Action and Design Research (PAR/PD) program for adolescent online safety called "Teenovate." Through synergistic efforts with the Teenovate program, Wisniewski and Lee will conduct PD workshops with thirty adolescents to evaluate and iterate on the design of the web-based experiential intervention program for cybergrooming. The workshop will be iterative in that teen participants will engage with the researchers twice - once to share their feedback and initial design ideas, then again once these designs have been incorporated into the final prototype. In some cases, if teens are highly motivated to join the project team on an ongoing basis to see their designs come to fruition, we will form a Youth Advisory Board (YAB) so that they can continuously provide their input. PD approaches are well-established within the HCI literature [147] and involve engaging directly with the people who will ultimately use a system in the design of that system. Like Wisniewski's ongoing PD studies, teens will first be trained to use industry-standard User Experience (UX) tools, such as Envision and Figma, to learn the skills to create high-fidelity interactive prototypes of their design ideas. Then, we will ask teen participants to complete the web-based cybergrooming experiential intervention, including tasks where they interact with the chatbot. We will solicit their feedback on the usability, validity, and authenticity of the training and chatbot responses, as well as ask their opinion on how the system could be redesigned to be engaging and effective. We will then ask teens to create mock-ups for how they would redesign or improve both the training module and interactive chatbot interface. This feedback will be qualitatively analyzed and used to iteratively improve the design of the experiential intervention program, as well as to refine the chatbots developed in Thrust AI.

System Design and Implementation. UCF PI Wisniewski and Co-PI Lee will jointly lead a team of

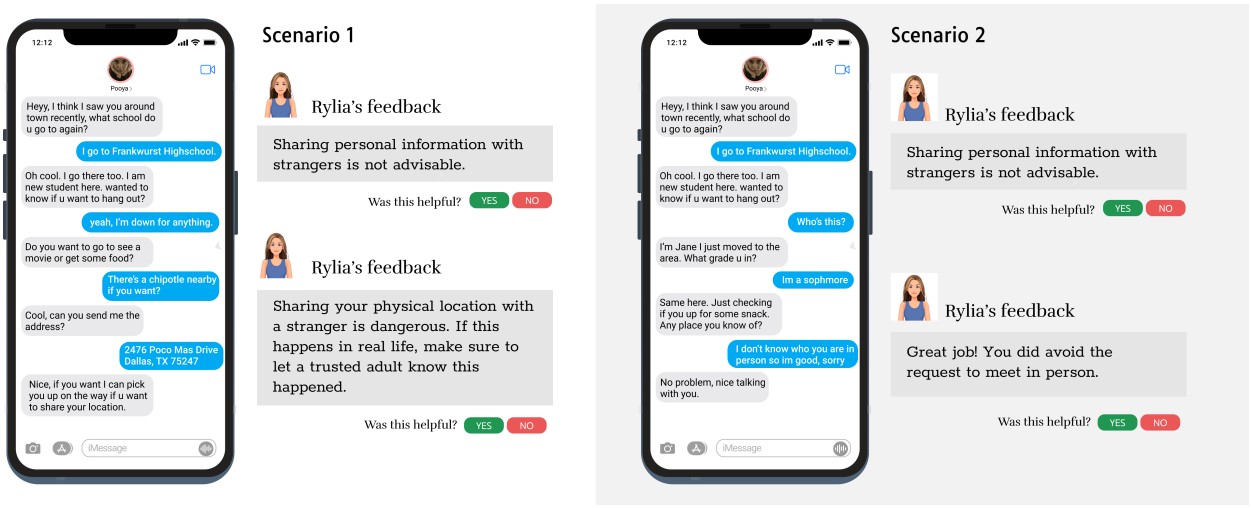

Figure 5: Prototypes of experiential intervention for enhancing resiliency to cybergrooming risks. Adolescents participants will interactively converse with the perpetrator chatbot (gray text bubbles) and will get post-scenario feedback on their responses (blue text bubbles). The participant on the left scenario showed high-level of vulnerability than one on the right one, receiving appropriate feedback from the program.

Computer Science Senior Design students in embedding the training module along with the interactive chatbot in a web-based interface to create a fully functional prototype of the cybergrooming experiential intervention. Figure 5 provides an example web-based chatbot interface with examples that would make a teen highly vulnerable (left) to protected (right) from cybergrooming advances. Based on the findings from the PD workshops and experts feedback, the conversations will be modified to keep the chatbot messages age-appropriate, so that teens will not be exposed to content that may be triggering. The developed intervention may employ various strategies – interactive scenarios with actionable feedback, gamifying the risk detection, and/or implementing in authentic context, such as social media or online games. We will deploy the program online for enabling large-scale data collection and evaluation in Thrust HD, as well as widespread dissemination once our proof-of-concept has been validated.

6.3 Thrust HD: Assessment of Adolescents' Risk Coping Skills Against Cybergrooming

Objective: Thrust HD will aim to assess adolescents' risk coping skills against cybergrooming with two tasks. First, we will assess the effectiveness of the proposed intervention program and identify vulnerability factors through a participatory community-based survey. Second, we will cross-validate the findings from survey/interview-based and NLP-based assessment of teens' vulnerability/resilience to cybergrooming.

Thrust HD: Research Questions

1. *Does the proposed chatbot-based experiential intervention program increase adolescents' awareness of cybergrooming victimization processes and their risk-coping skills against them?*
2. *Which biopsychosocial risk and resilient factors are associated with adolescents' resistance and receptivity to experiential intervention for cybergrooming?*
3. *To what extent can the cross-validation using both social science and text-mining methodologies (e.g., surveys and NLP) improve the validity of the vulnerability analysis of adolescents to cybergrooming?*

This thrust will assess the effectiveness of the proposed chatbot-based experiential intervention program in Thrust AI for enhancing adolescents' awareness to cybergrooming and developing their effective risk coping skills to handle it. To this end, we will investigate the following hypotheses:

- **H1**: Adolescents' frequent Internet use will directly affect increasing their vulnerability to cybergrooming.
- **H2**: Adolescents' vulnerability to cybergrooming will substantially increase if they have risk factors, such as sensation seeking and impulsive traits, low self-esteem, and adjustment problems at home and school.
- **H3**: Vulnerability to cybergrooming will not be observed among adolescents whose resilience is enhanced through the proposed experiential intervention program.

We will measure adolescents' vulnerability in terms of the extent of their engagement in the conversations with a perpetrator by providing their private information and/or committing an in-person meeting with the perpetrator. Figure 6 summarizes the three key hypotheses to be studied in this thrust. Specifically, path H1 describes how the frequent use of Internet can affect adolescents' vulnerability to cybergrooming victimization. Path H2 will investigate what biopsychosocial vulnerability factors can introduce adolescents' vulnerability to cybergrooming. Lastly, path H3 will assess the effectiveness of the proposed chatbot-based experiential intervention program on building adolescents' resilience to cybergrooming.

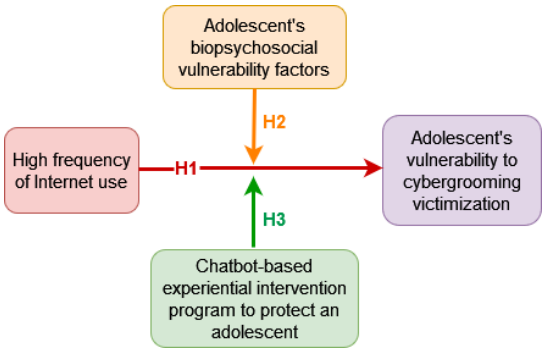

Figure 6: A Theoretical Model for Thrust HD.

6.3.1 Task HD.A: Assessment of the Program Effectiveness and Identification of Vulnerability and Resilience Factors through a Participatory Community-Based Survey

Survey Administration. To understand the roles of risk and resilience factors in increasing vulnerability to victimization as well as interacting with the proposed experiential intervention program's effectiveness, a survey will be administered online. More specifically, we will employ pre/post-test based evaluation design, which is defined as a before and after the assessment to measure whether the expected changes take place in the adolescent participants' understanding and awareness of the victimization processes. During the pre-test, adolescents will be asked to answer a set of questions measuring the following biopsychosocial factors: age, gender, average daily time spent on the Internet, disability, self-esteem, risk-taking behaviors, parent-child relationship, school experiences, and previous experience of abuse and victimization.

Statistical Analyses. To quantify changes in adolescents' awareness of cybergrooming victimization processes as a function of the proposed experiential intervention program, we will perform average-based change statistics (ABC), such as Cohen's d or Hays' ω^2 as well as individual-based change statistics (IBC) [148]. The ABC will provide information on group-based systematic changes, which will help the researchers evaluate the effectiveness of the proposed experiential intervention program. The IBC will provide insights on individual differences in the systematic changes in the level of awareness of cybergrooming victimization processes, which will set a platform for inferential statistical analyses focusing on identifying factors that either increase or decrease the effectiveness of the program at individual level. Regression and structural equation modeling [149] will be the main statistical analyses to utilize.

6.3.2 Task HD.B: Cross-Validation of Adolescents' Vulnerability and Resilience to Cybergrooming

Evaluating the Chatbot-based Cybergrooming Experiential Learning Program. To assess the effectiveness of the chatbot-based experiential intervention program to prevent cybergrooming in Thrust AI in terms of the likelihood of cybergrooming, we will execute pre-testing, and post-testing. In this research design, the same assessment measures will be administered both before and after participants have completed the proposed program. The same measures will be used to determine if statistically significant changes are detected as a function of the proposed program's effectiveness. For the pre-testing, the measurement of risk coping skills to cybergrooming will be presented at the end of the online survey described above under Task HD.A. The post-testing will be performed as an exit survey after participants have completed the proposed experiential intervention program developed in Thrust AI.

NLP-based Assessment of Adolescents' Vulnerability and Resilience to Cybergrooming. According to the extensive review in [150], young people's key vulnerability factors to cybergrooming are identified at the *individual*, *family*, *community*, and *cultural* (i.e., country-based) levels. We limit our scope to analyzing the vulnerability aspects of adolescents in the U.S. and only consider the vulnerabilities at the first three.

We will leverage the existing NLP tools, such as LIWC [151] and Receptive API [152], to measure those vulnerabilities from the conversations between a perpetrator chatbot and an adolescent user collected in Thrust HCI. Based on the resilience and vulnerability factors identified in Task HD.A, we will measure their extents based on the languages the adolescents use where they can be categorized as vulnerabilities at individual, family, and community levels, as described in Table 4.

These traits will be measured based on categories and feature measurements available in text mining tools [151, 152]. To understand the degree of adolescents' vulnerability to cybergrooming, we describe example key attributes to be measured using text-mining tools, as shown in Table 4. Based on the scores of each attribute measured by various features available in NLP tools (e.g., LIWC's categories and features), we will investigate the correlations between vulnerability factors and being a potential victim to cybergrooming. We have conducted preliminary research in [90] using the PJ datasets. However, we did not conduct cross-validation using real adolescents' datasets and survey-based analysis in [90]. This proposed thrust will fill the gaps. Since the PJ dataset provides the conversation languages by professionally trained volunteers playing potential victims, it may not provide real teens' language behaviors under cybergrooming situations. To mitigate this, we will also collect real conversation datasets from real teens in Thrust HCI, and train our models in Thrust AI and analyze the vulnerability using the conversation datasets collected in Thrust HCI.

Table 4: MEASUREMENTS OF EXAMPLE VULNERABILITY FACTORS TO CYBERGROOMING

Type	Key Dimensions	Key Attributes
Individual	Identity traits	Sexual orientation, self-confidence, self-esteem, sensation seeking, self-perception
	Personality traits	Positive emotion, negative emotion*, anxiety*, anger*, sadness*, risk taking
	Cognitive traits	Risk taking behavior, cognitive ability, susceptibility to persuasion*
Family	Family relationships	Relationships with family members or significant others
	SES	Family's social and economic status
Comm.	Social support	Friends, peer support
	Community support	Quality of living environments; school support

Follow-up Interviews. Adolescent participants who complete the pre-post evaluation of our experiential learning program will be invited to take part in a follow-up interview to share their experience and opinions. Interviews will first probe whether teens found the program beneficial and improvements that could be made. Then, we will further probe based on the teens' individual, family, and community dimensions and attributes to assess whether patterns emerge based on the individual characteristics of the teens, which will establish the generalizability and applicability of our program across diverse backgrounds.

7 Evaluation & Experiment Plan

Experiment Settings and Setup: (1) **Thrust AI**: For possible noisy languages used in online chatting conversations, we will remove the noises (e.g., Emojis, Mentions, URLs, or Hashtags) by a regular expression-based Python library *Preprocessor* and add essential spaces to non-segmented words from informal languages using *wordsegment* library. We will also employ MoNoise [153], a state-of-the-art lexical normalization model, to normalize lexical slangs in datasets. The entailment analysis framework designed to extract categorical grooming acts of perpetrator and victims will be based on the state-of-the-art pretrained language models, such as BERT (Bidirectional Encoder Representations from Transformers) [117] or T5 (Text-to-Text Transfer Transformer) model [154], and optimized on the large-scale natural language inference datasets as detailed in the data management plan. For knowledge grounded conversation generation, we will take the whole WikiData [155] and ConceptNet [156] as the source of world fact and commonsense knowledge. To predict a grooming stage of each dialogue utterance, we will apply TextCNN [157] and optimize it on the annotated labels from the PJ dataset. The chatbots will be first pre-trained on large-scale casual chat datasets listed in the data management plan, and fine-tuned on the domain-specific PJ dataset and collected datasets from Thrust HCI. (2) **Thrust HCI**: This thrust will be inherently evaluative in nature by explicitly asking key stakeholders (e.g., parents, teens, and domain experts) to inform the design and development of both the training dataset deriving the chatbots and the web-based system. This will culminate in the experiential intervention program for cybergrooming as the final end product of this proposal. In Thrust HCI.A, evaluative measures will be both quantitative and qualitative in nature as we will ask adult-adolescent crowdworkers to vote on the quality of different text-based response and to contribute to the dataset in their own language. Thrust HCI.B will be qualitative and generative by asking teens to give their feedback on our initial system as well as creating designs of their own that will be incorporated into training the proposed experiential intervention program further. Thrust HCI will be user-centered to ensure that our participants and stakeholders drive the final design of the system and training modules, rather than the research team; and (3) **Thrust HD**: To conduct human experiments, we will leverage the facilities of the ISERC that allows research and education space for complex scenarios for a full spectrum of experiments,

as described in detail in in the FEOR document. We will also leverage various NLP tools (e.g., LIWC, Receptive API, or SEANCE) to conduct social-psychological analysis from lexical and behavioral features based on conversation datasets. For statistical analysis, we will use the average-based change statistics (ABC), such as Cohen's d or Hays' ω^2 as well as individual-based change statistics (IBC).

Metrics: (1) **Thrust AI**: We will evaluate the fluency of the proposed chatbot's conversations based on referenced metrics (i.e., BLEU [158], ROUGE [159] and BERTScores [160]) and unreferenced metrics (i.e., perplexity and MaUde scores [161]). In addition, we will measure the degrees of vulnerability factors identified from the adolescent participants' conversation languages with the chatbot. We will also validate the performance of the proposed DRL-based, strategic conversation models based on the chatbot's converged reward and speed of the reward convergence. (2) **Thrust HCI**: In Thrust HCI.A, statistical calculations based on the weighted averages of crowdworker votes will be used to rank the quality and ecological validity of different responses. For Thrust HCI.B, qualitative feedback from adolescent users of the system as well as standard Likert-scale measures of usability and preference will be leveraged to evaluate user perceptions of the experiential intervention for cybergrooming. (3) **Thrust HD**: We will measure the levels of vulnerability and resilience based on both NLP tools and surveys. By combining the results from the NLP-based vulnerability scores and the survey-based answers obtained from Thrust HD.A, we will investigate correlations between key factors of interest along with their statistical significance levels (e.g., p-value).

Datasets: We described the details of datasets [108, 162, 163, 164, 165, 166, 167, 124, 154, 168, 9, 169, 170] to be used in this research in the submitted data management plan.

8 Ethical Considerations

We will assure that the research outcome from this proposed work will not be used in negative fashion, e.g., using the chatbot framework as a tool for automating online sexual exploitation or abuse of children. Therefore, we will selectively make our research outcome available to avoid any malicious application of the proposed chatbot framework. We provide some example scenarios where the chatbot framework should or should not be used: (1) **Should-Do:** (a) Educational parties using the chatbot framework to develop curricula to educate adolescents about how to respond to online abusive messages and cope with cybergrooming; (b) sharing datasets selectively with research groups wanting to use it for the purpose of protecting adolescents from online risks; and (c) parents wishing to learn grooming conversations to educate their children to build resistance and resilience against the potential risk of cybergrooming. (2) **Should-Not-Do:** (a) To mitigate any possible misuse of sensitive or inappropriate languages used by the chatbots, we will eliminate them using linguistic resources, such as the profane lexicons [171], to replace offensive/profane words in the training dataset with moderate ones and balance between simulating a realistic cybergrooming scenario and avoiding any potential ethical issues; (b) We will only make our source code and model to be accessible to parties providing a clear research goal and identity; (c) We will design approaches to monitor the language generated by a perpetrator chatbot and a user and make the chatbot close a session with a warning message whenever offensive/profane language is detected. This is to prevent potential misuse of this proposed technology by malicious users (e.g., a potential perpetrator); (d) We will make all the conversational data collected encrypted and stored under the regulations and standards stated in the legal frameworks, such as the General Data Protection Regulation [172] and strictly follow our submitted data management plan; and (e) For all human subjects research, we will follow best practices established from PI Wisniewski's prior work with minors [173, 174, 175] and IRB guidelines when handling potentially privacy invasive, illegal, risky, and/or otherwise sensitive topics among adolescents.

9 Broader Impacts

Societal Impact: We will provide a multidisciplinary research solution to solve a serious social problem, online child sexual exploitation and abuse through 'cybergrooming.' Based on the key findings of this research, we will develop and evaluate a cybergrooming experiential learning program for adolescents. By combining the advanced features of AI technologies with the social science research methodologies, this research will provide novel cross-validation of the proposed approach that can ensure efficient and effective risk and vulnerability assessment for protecting children from cybergrooming and further building their resilience to cybergrooming. Our ultimate goal is to deploy a robust and effective web-based cybergrooming intervention program that can be provided at no costs to English-speaking families world-wide. By giv-

ing parents and adolescents a low-risk way to discuss and develop risk-coping strategies against realistic cybergrooming scenarios, the broader impact is to prevent cybervictimization before it occurs.

Contribution to Curriculum Development: We will develop presentations, case studies, and exercises that incorporate the research problems as well as the results of our research into existing courses (e.g., VT PI Cho –'Network Security,' Co-PI Huang –'Natural Language Processing,' Co-PI Wisniewski – 'Participatory Design for Empowering Vulnerable Populations,' Co-PI Lee –'CSCW and Social Computing,' Co-PI Kim – 'Research Methods in Human Development') to encourage students to become familiar with the relevant problems, processes, models, and applications for AI-mediated intervention addressing online risks.

Minorities and Outreach: We will encourage the participation of students from underrepresented groups through the following programs: VT's programs including Multicultural Academic Opportunities Program (MOAP) [176], the Center for Enhancement of Engineering Diversity (CEED) [177], the McNair Program [178], and the National Center for Women and VT's Information Technology (NCWIT) [179]. UCF programs include ACM-W, McKnight, NSF-funded S-STEM EXCEL, Flit-PATH, Flit-GAP, and the College of Engineering and Computer Science's Office of Diversity and Inclusion. All (Co)PIs have worked with more than 10 female PhD students during the past 5 years.

Transdisciplinary Research and Education: The proposed research is highly transdisciplinary across machine/deep learning, natural language processing, human-computer interaction, cybersecurity, and human development and sciences, and encouraging broader, newer, diverse approaches of problem-solving skills. To widespread the impact of this transdisciplinary research, we will leverage undergraduate programs providing a major in 'Interdisciplinary Studies' at VT [180] to provide an opportunity for students to explore interdisciplinary learning by providing seminar talks, workshops, short courses, or research projects. At UCF, Co-PI Wisniewski will engage students across the Computer Science, Information Technology, Psychology Human Factors, Modelling, Simulation, and Training, and Criminal Justice programs.

Dissemination of Research Outcomes: We will make all our research findings available to the public by publishing them in peer-reviewed conferences or journals. We will follow a privacy policy for all datasets used by anoymizing private information. For the source code and software, we will make them available to the parties providing their clear identity and research purposes to avoid any ethical issues as in Section 8. We provided the details of disseminating research outcomes in the submitted data management plan.

Broadening Participation in Computing (BPC): We proposed a BPC program to work with students from underrepresented groups through VT's CEED program, called BLAST (Building Leaders for Advancing Science and Technology). We detailed our proposed BPC program in the submitted BPC document.

10 Results from Prior NSF Support

(Co)PIs are working on multiple NSF projects. However, due to space constraint, we only list the projects that can be best leveraged in this project. **Lead PI Cho** (PI, 2107450, 10/1/2021 – 09/31/2025, $500,000), "III: Medium: Collaborative Research: MUDL: Multidimensional Uncertainty-Aware Deep Learning Framework." **Intellectual Merit**: Developing multidimensional uncertainty-aware decision making framework developed by bridging a belief model and deep learning [181, 71, 182]. **Broader Impact**: Introducing a high-impact on DL-based applications in terms of modeling, prediction, and interpretation of different types of uncertainties and their impact on decision making performance. **UCF PI Wisniewski** (PI, 1827700, 09/15/2018 - 02/28/2023, $810,550): IIP: PFI-RP: Using a Human-Centered Approach to Improve Algorithms for Detecting Adolescent Online Risks." **Intellectual Merit:** Leveraging human-centered approaches to design, develop, and evaluate patterns, end-user technologies, and algorithms for promoting online privacy, security, and safety [57, 57, 104, 61, 59, 141, 143]. **Broader Impacts:** Promoting the online safety, security, and privacy of adolescents. **Co-PI Lee** (PI, 2119011, 2021/10-2024/9, $849,999): "RETTL: Facilitating socially constructed learning through a shared, mobile-based virtual reality platform in informal learning settings." **Intellectual Merit**: Experimenting and validating a group-based VR platform using mobile devices, where learners are exposed to STEM content in a socially-connected way. **Broader Impacts**: Enabling immersive, inclusive, and social VR experiences including those who do not wear VR headsets due to psychological and physical issues. **Co-PIs Huang and Kim**: No NSF support in recent five years.

References

[1] NCMEC. (2021) National center for missing and exploited children. [Online]. Available: https://www.missingkids.org/home

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

[3] S. Marchenko. (2017) Web of darkness: Groomed, manipulated, coerced, and abused in minutes. [Online]. Available: https://www.biometrica.com/icmec-online-grooming/

[4] J. Davidson, J. Grove-Hills, A. Bifulco, P. Gottschalk, V. Caretti, T. Pham, and S. Webster, "European online grooming project: Summary of the project findings," *European Online Grooming Project. Retrieved February*, vol. 23, p. 2012, 2011.

[5] Y. Zheng, G. Chen, M. Huang, S. Liu, and X. Zhu, "Personalized dialogue generation with diversified traits," *arXiv preprint arXiv:1901.09672*, 2019.

[6] H. Song, W.-N. Zhang, Y. Cui, D. Wang, and T. Liu, "Exploiting persona information for diverse generation of conversational responses," *arXiv preprint arXiv:1905.12188*, 2019.

[7] H. Song, W.-N. Zhang, J. Hu, and T. Liu, "Generating persona consistent dialogues by exploiting natural language inference," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, 2020, pp. 8878–8885.

[8] Q. Liu, Y. Chen, B. Chen, J.-G. Lou, Z. Chen, B. Zhou, and D. Zhang, "You impress me: Dialogue generation via mutual persona perception," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 1417–1427.

[9] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston, "Personalizing dialogue agents: I have a dog, do you have pets too?" in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 2204–2213.

[10] J. Li, M. Galley, C. Brockett, G. Spithourakis, J. Gao, and W. B. Dolan, "A persona-based neural conversation model," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, pp. 994–1003.

[11] X. Kong, B. Li, G. Neubig, E. Hovy, and Y. Yang, "An adversarial approach to high-quality, sentiment-controlled neural dialogue generation," *arXiv preprint arXiv:1901.07129*, 2019.

[12] M. Firdaus, H. Chauhan, A. Ekbal, and P. Bhattacharyya, "More the merrier: Towards multi-emotion and intensity controllable response generation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 14, 2021, pp. 12 821–12 829.

[13] Q. Li, H. Chen, Z. Ren, Z. Chen, Z. Tu, and J. Ma, "Empgan: Multi-resolution interactive empathetic dialogue generation," *arXiv e-prints*, pp. arXiv–1911, 2019.

[14] Z. Deng, H. Lin, W. Huang, R. Lan, and X. Luo, "Emotional dialogue generation based on conditional variational autoencoder and dual emotion framework," *Wireless Communications and Mobile Computing*, vol. 2020, 2020.

[15] H. Rashkin, E. M. Smith, M. Li, and Y.-L. Boureau, "Towards empathetic open-domain conversation models: A new benchmark and dataset," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 5370–5381.

[16] L. Luo, W. Huang, Q. Zeng, Z. Nie, and X. Sun, "Learning personalized end-to-end goal-oriented dialog," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 6794–6801.

[17] A. Bordes, Y.-L. Boureau, and J. Weston, "Learning end-to-end goal-oriented dialog," *arXiv preprint arXiv:1605.07683*, 2016.

[18] C. Xu, W. Wu, and Y. Wu, "Towards explainable and controllable open domain dialogue generation with dialogue acts," *arXiv preprint arXiv:1807.07255*, 2018.

[19] R. Lian, M. Xie, F. Wang, J. Peng, and H. Wu, "Learning to select knowledge for response generation in dialog systems," in *IJCAI International Joint Conference on Artificial Intelligence*, 2019, p. 5081.

[20] W. Zheng, N. Milic-Frayling, and K. Zhou, "Approximation of response knowledge retrieval in knowledge-grounded dialogue generation," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, 2020, pp. 3581–3591.

[21] E. Dinan, S. Roller, K. Shuster, A. Fan, M. Auli, and J. Weston, "Wizard of wikipedia: Knowledge-powered conversational agents," in *International Conference on Learning Representations*, 2018.

[22] C. Zhang, H. Wang, F. Jiang, and H. Yin, "Adapting to context-aware knowledge in natural conversation for multi-turn response selection," in *Proceedings of the Web Conference 2021*, 2021, pp. 1990–2001.

[23] X. Lin, W. Jian, J. He, T. Wang, and W. Chu, "Generating informative conversational response using recurrent knowledge-interaction and knowledge-copy," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 41–52.

[24] J. Wang, J. Liu, W. Bi, X. Liu, K. He, R. Xu, and M. Yang, "Improving knowledge-aware dialogue generation via knowledge base question answering," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, 2020, pp. 9169–9176.

[25] T. Young, E. Cambria, I. Chaturvedi, H. Zhou, S. Biswas, and M. Huang, "Augmenting end-to-end dialogue systems with commonsense knowledge," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[26] H. Zhang, Z. Liu, C. Xiong, and Z. Liu, "Grounded conversation generation as guided traverses in commonsense knowledge graphs," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 2031–2043.

[27] H. Zhou, T. Young, M. Huang, H. Zhao, J. Xu, and X. Zhu, "Commonsense knowledge aware conversation generation with graph attention," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018, pp. 4623–4629.

[28] P. Zhou, K. Gopalakrishnan, B. Hedayatnia, S. Kim, J. Pujara, X. Ren, Y. Liu, and D. Hakkani-Tur, "Commonsense-focused dialogues for response generation: An empirical study," *arXiv preprint arXiv:2109.06427*, 2021.

[29] B. P. Majumder, H. Jhamtani, T. Berg-Kirkpatrick, and J. McAuley, "Like hiking? you probably enjoy nature: Persona-grounded dialog with commonsense expansions," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 9194–9206.

[30] E. Levin, R. Pieraccini, and W. Eckert, "Learning dialogue strategies within the markov decision process framework," in *1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings*, 1997, pp. 72–79.

[31] ——, "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)*, vol. 1, 1998, pp. 201–204 vol.1.

[32] S. Singh, M. Kearns, D. Litman, and M. Walker, "Reinforcement learning for spoken dialogue systems," *Advances in Neural Information Processing Systems (NIPS)*, vol. 12, pp. 956–962, 1999.

[33] B. Rofi'ah, H. Fakhrurroja, and C. Machbub, "Dialogue management using reinforcement learning." *Telkomnika*, vol. 19, no. 3, 2021.

[34] N. Roy, J. Pineau, and S. Thrun, "Spoken dialogue management using probabilistic reasoning," in *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, 2000, pp. 93–100.

[35] L. Daubigney, M. Geist, S. Chandramohan, and O. Pietquin, "A comprehensive reinforcement learning framework for dialogue management optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 8, pp. 891–902, 2012.

[36] R. Lan, J. Wang, W. Huang, Z. Deng, X. Sun, Z. Chen, and X. Luo, "Chinese emotional dialogue response generation via reinforcement learning," *ACM Transactions on Internet Technology (TOIT)*, vol. 21, no. 4, pp. 1–17, 2021.

[37] R. Zhang, Z. Wang, M. Zheng, Y. Zhao, and Z. Huang, "Emotion-sensitive deep dyna-q learning for task-completion dialogue policy learning," *Neurocomputing*, vol. 459, pp. 122–130, 2021.

[38] J. Li, W. Monroe, A. Ritter, D. Jurafsky, M. Galley, and J. Gao, "Deep reinforcement learning for dialogue generation," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, Nov. 2016, pp. 1192–1202. [Online]. Available: https://aclanthology.org/D16-1127

[39] B. Peng, X. Li, J. Gao, J. Liu, and K.-F. Wong, "Deep Dyna-Q: Integrating planning for task-completion dialogue policy learning," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 01 2018, pp. 2182–2192.

[40] P. Dekker and R. V. Rooy, "Bi-directional optimality theory: An application of game theory," *Journal of Semantics*, vol. 17, pp. 217–242, 2000.

[41] I. Lewin and M. Lane, "A formal model of conversational game theory," in *Proceedings of the 4th workshop on the semantics and pragmatics of dialogue (Gotalog)*, vol. 69. Citeseer, 2000.

[42] A. Cadilhac, N. Asher, F. Benamara, and A. Lascarides, "Grounding strategic conversation: Using negotiation dialogues to predict trades in a win-lose game," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, pp. 357–368.

[43] M. Kacprzak, M. Dziubiński, and K. Budzynska, "Strategies in dialogues: A game-theoretic approach," in *Computational Models of Argument*. IOS Press, 2014, pp. 333–344.

[44] M. Barlier, J. Perolat, R. Laroche, and O. Pietquin, "Human-machine dialogue as a stochastic game," in *16th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL 2015)*, 2015.

[45] R. Laroche and A. Genevay, "The negotiation dialogue game," in *Dialogues with Social Robots*. Springer, 2017, pp. 403–410.

[46] K. Van Deemter, "What game theory can do for NLG: The case of vague language (invited talk)," in *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)*, 2009, pp. 154–161.

[47] A. Razi, S. Kim, A. Alsoubai, G. Stringhini, T. Solorio, M. De Choudhury, and P. J. Wisniewski, "A human-centered systematic literature review of the computational approaches for online sexual risk detection," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW2, oct 2021. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/3479609

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

[49] F. Muñoz, G. Isaza, and L. Castillo, "Smartsec4cop: Smart cyber-grooming detection using natural language processing and convolutional neural networks," in *International Symposium on Distributed Computing and Artificial Intelligence*. Springer, 2020, pp. 11–20.

[50] M. Ashcroft, L. Kaati, and M. Meyer, "A step towards detecting online grooming–identifying adults pretending to be children," in *2015 European Intelligence and Security Informatics Conference*. IEEE, 2015, pp. 98–104.

[51] D. Bogdanova, P. Rosso, and T. Solorio, "Exploring high-level features for detecting cyberpedophilia," *Computer speech & language*, vol. 28, no. 1, pp. 108–120, 2014.

[52] A. E. Cano, M. Fernandez, and H. Alani, "Detecting child grooming behaviour patterns on social media," in *International conference on social informatics*. Springer, 2014, pp. 412–427.

[53] F. E. Gunawan, L. Ashianti, S. Candra, and B. Soewito, "Detecting online child grooming conversation," in *11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, 2016, pp. 1–6.

[54] A. Gupta, P. Kumaraguru, and A. Sureka, "Characterizing pedophile conversations on the internet using online grooming," *arXiv preprint arXiv:1208.4324*, 2012.

[55] J. Tomljanovic, L. Zuanovic, and T. Šebrek, "Sexual predator identification using word2vec features," *Text Analysis and Retrieval 2016 Course Project Reports*, p. 70, 2016.

[56] M. Charalambous, P. Papagiannis, A. Papasavva, P. Leonidou, R. Constaninou, L. Terzidou, T. Christophides, P. Nicolaou, O. Theofanis, G. Kalatzantonakis, and M. Sirivianos, "A privacy-preserving architecture for the protection of adolescents in online social networks," 2020.

[57] P. Wisniewski, A. K. Ghosh, H. Xu, M. B. Rosson, and J. M. Carroll, "Parental control vs. teen self-regulation: Is there a middle ground for mobile online safety?" in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 51–69. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/2998181.2998352

[58] P. Wisniewski, H. Jia, N. Wang, S. Zheng, H. Xu, M. B. Rosson, and J. M. Carroll, "Resilience mitigates the negative effects of adolescent internet addiction and online risk exposure," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 4029–4038. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/2702123.2702240

[59] A. K. Ghosh, K. Badillo-Urquiola, S. Guha, J. J. LaViola Jr, and P. J. Wisniewski, *Safety vs. Surveillance: What Children Have to Say about Mobile Apps for Parental Control*. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–14. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/3173574.3173698

[60] H. Jia, P. J. Wisniewski, H. Xu, M. B. Rosson, and J. M. Carroll, "Risk-taking as a learning process for shaping teen's online information privacy behaviors," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 583–599. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/2675133.2675287

[61] A. K. Ghosh, K. Badillo-Urquiola, M. B. Rosson, H. Xu, J. M. Carroll, and P. J. Wisniewski, *A Matter of Control or Safety? Examining Parental Use of Technical Monitoring Apps on Teens' Mobile Devices*. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–14. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/3173574.3173768

[62] T. L. Rutkowski, H. Hartikainen, K. E. Richards, and P. J. Wisniewski, "Family communication: Examining the differing perceptions of parents and teens regarding online safety communication," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW2, oct 2021. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/3479517

[63] S. Wachs, G. K. Jiskrova, A. T. Vazsonyi, K. D. Wolf, and M. Junger, "A cross-national study of direct and indirect effects of cyberbullying on cybergrooming victimization via self-esteem," *Psicologia Educativa*, vol. 22, no. 1, pp. 61–70, 2016.

[64] S. E. Baumgartner, P. M. Valkenburg, and J. Peter, "Unwanted online sexual solicitation and risky sexual online behavior across the lifespan," *Journal of Applied Developmental Psychology*, vol. 31, no. 6, pp. 439–447, 2010.

[65] P. De Santisteban and M. Gámez-Guadix, "Prevalence and risk factors among minors for online sexual solicitations and interactions with adults," *The Journal of Sex Research*, vol. 55, no. 7, pp. 939–950, 2018.

[66] J. J. Laird, B. Klettke, K. Hall, E. Clancy, and D. Hallford, "Demographic and psychosocial factors associated with child sexual exploitation: A systematic review and meta-analysis," *JAMA Network Open*, vol. 3, no. 9, pp. e2 017 682–e2 017 682, 2020.

[67] G. M. Winters, L. E. Kaylor, and E. L. Jeglic, "Sexual offenders contacting children online: an examination of transcripts of sexual grooming," *Journal of Sexual Aggression*, vol. 23, no. 1, pp. 62–76, 2017.

[68] S. Wachs, K. D. Wolf, and C.-C. Pan, "Cybergrooming: Risk factors, coping strategies and associations with cyberbullying," *Psicothema*, pp. 628–633, 2012.

[69] S. Yoon, J. Cho, D. S. Kim, T. J. Moore, F. F. Nelson, H. Lim, N. Leslie, and C. Kamhoua, "Moving target defense for in-vehicle software-defined networking: IP shuffling in network slicing with multiagent deep reinforcement learning," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II*, T. Pham, L. Solomon, and K. Rainey, Eds., vol. 11413, International Society for Optics and Photonics. SPIE, 2020, pp. 341 – 350. [Online]. Available: https://doi.org/10.1117/12.2557850

[70] S. Yoon, J.-H. Cho, G. Dixit, and I.-R. Chen, "Resource-aware intrusion response based on deep reinforcement learning for software-defined Internet-of-Battle-Things," *Game Theory and Machine Learning for Cyber Security*, pp. 389–409, 2021.

[71] X. Zhao, S. Hu, J. Cho, and F. Chen, "Uncertainty-based decision making using deep reinforcement learning," in *22nd IEEE FUSION 2019*, Ottawa, CA, 2–5 Jul. 2019.

[72] Q. Zhang, J. Cho, T. J. Moore, and F. F. Nelson, "DREVAN: Deep reinforcement learning-based vulnerability-aware network adaptations for resilient networks," in *accepted to The 2021 IEEE Conference on Communications and Network Security (CNS 2021)*, 2021.

[73] J.-H. Cho, M. Zhu, and M. Singh, *Modeling and Analysis of Deception Games Based on Hypergame Theory*. Cham: Springer International Publishing, 2019, pp. 49–74. [Online]. Available: https://doi.org/10.1007/978-3-030-02110-8_4

[74] Z. Wan, J.-H. Cho, M. Zhu, A. H. Anwar, C. Kamhoua, and M. P. Singh, "Foureye: Defensive deception against advanced persistent threats via hypergame theory," *IEEE Transactions on Network and Service Management*, pp. 1–1, 2021.

[75] X. Feng, L. Huang, D. Tang, H. Ji, B. Qin, and T. Liu, "A language-independent neural network for event detection," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2016, pp. 66–71.

[76] L. Huang, T. Cassidy, X. Feng, H. Ji, C. Voss, J. Han, and A. Sil, "Liberal event extraction and event schema induction," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, pp. 258–268.

[77] L. Huang, J. May, X. Pan, H. Ji, X. Ren, J. Han, L. Zhao, and J. A. Hendler, "Liberal entity extraction: Rapid construction of fine-grained entity typing systems," *Big Data*, vol. 5, no. 1, pp. 19–31, 2017.

[78] L. Huang, H. Ji, K. Cho, I. Dagan, S. Riedel, and C. Voss, "Zero-shot transfer learning for event extraction," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 2160–2170.

[79] L. Huang and H. Ji, "Semi-supervised new event type induction and event detection," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 718–724.

[80] Y. Cao, L. Huang, H. Ji, X. Chen, and J. Li, "Bridge text and knowledge by learning multi-prototype entity mention embedding," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2017, pp. 1623–1633.

[81] D. Yu, L. Huang, and H. Ji, "Open relation extraction and grounding," in *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2017, pp. 854–864.

[82] D. Lu, S. Whitehead, L. Huang, H. Ji, and S.-F. Chang, "Entity-aware image caption generation," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 4013–4023.

[83] Q. Wang, X. Pan, L. Huang, B. Zhang, Z. Jiang, H. Ji, and K. Knight, "Describing a knowledge base," in *Proceedings of the 11th International Conference on Natural Language Generation*, 2018, pp. 10–21.

[84] Q. Wang, L. Huang, Z. Jiang, K. Knight, H. Ji, M. Bansal, and Y. Luan, "Paperrobot: Incremental draft generation of scientific ideas," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 1980–1991.

[85] Q. Wang, Z. Zhou, L. Huang, S. Whitehead, B. Zhang, H. Ji, and K. Knight, "Paper abstract writing through editing mechanism," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2018, pp. 260–265.

[86] Q. Wang, Q. Zeng, L. Huang, K. Knight, H. Ji, and N. F. Rajani, "Reviewrobot: Explainable paper review generation based on knowledge synthesis," in *Proceedings of the 13th International Conference on Natural Language Generation*, 2020, pp. 384–397.

[87] S. Zhang, L. Huang, and Q. Ning, "Extracting temporal event relation with syntactic-guided temporal graph transformer," *arXiv preprint arXiv:2104.09570*, 2021.

[88] P. Wang, Z. Guo, L. Huang, and J. Cho, "SERI: Generative chatbot framework for cybergrooming prevention," in *The First Workshop on Evaluations and Assessments of Neural Conversation Systems (EANCS)*. The Association for Computational Linguistics (ACL), 2021.

[89] ——, "Deep reinforcement learning-based authentic dialogue generation for a cybergrooming prevention program," in *The 60th Conference of Association of Computational Linguistics (ACL)*, 2021, under review.

[90] Z. Guo, P. Wang, J.-H. Cho, L. Huang, and K. J. Kim, "Text mining-based social-psychological vulnerability analysis of potential victims to cybergrooming: Insights and lessons learned," *IEEE Intelligent Systems*, 2021, under review.

[91] P. Wisniewski, H. Xu, M. B. Rosson, D. F. Perkins, and J. M. Carroll, "Dear diary: Teens reflect on their weekly online risk experiences," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 3919–3930. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/2858036.2858317

[92] S. W. Lee, A. Willette, D. Koutra, and W. S. Lasecki, "The effect of social interaction on facilitating audience participation in a live music performance," in *Proceedings of the 2019 on Creativity and Cognition*, ser. C&C '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 108–120. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/3325480.3325509

[93] S. W. Lee, Y. Zhang, I. Wong, Y. Y., S. O'Keefe, and W. Lasecki, "SketchExpress: Remixing animations for more effective crowd-powered prototyping of interactive interfaces," in *Proceedings of the ACM Symposium on User Interface Software and Technology*, ser. UIST. ACM, 2017. [Online]. Available: https://doi.org/10.1145/3126594.3126595

[94] S. W. Lee and G. Essl, "Web-based temporal typography for musical expression and performance." in *NIME*. Citeseer, 2015, pp. 65–69.

[95] ——, "Live writing: Asynchronous playback of live coding and writing," in *Proceedings of the International Conference on Live Coding*, 2015.

[96] S. W. Lee and J. Freeman, "echobo: A mobile music instrument designed for audience to play," *Ann Arbor*, vol. 1001, pp. 48 109–2121, 2013.

[97] S. W. Lee, A. D. de Carvalho Jr, and G. Essl, "Crowd in c [loud]: Audience participation music with online dating metaphor using cloud service," 2016.

[98] S. W. Lee, R. Krosnick, S. Y. Park, B. Keelean, S. Vaidya, S. D. O'Keefe, and W. S. Lasecki, "Exploring real-time collaboration in crowd-powered systems through a UI design tool," *Proc. ACM Hum.-Comput. Interact.*, vol. 2, no. CSCW, pp. 104:1–104:23, Nov. 2018. [Online]. Available: http://doi.acm.org/10.1145/3274373

[99] Y. Chen, S. W. Lee, Y. Xie, Y. Yang, W. S. Lasecki, and S. Oney, "Codeon: On-demand software development assistance," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 6220–6231. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/3025453.3025972

[100] M. M. Bhuiyan, H. Whitley, M. Horning, S. W. Lee, and T. Mitra, "Designing transparency cues in online news platforms to promote trust: Journalists' & consumers' perspectives," vol. 5, no. CSCW2, Oct. 2021. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/3479539

[101] Y. Chen, J. Herskovitz, G. Matute, A. Wang, S. W. Lee, W. S. Lasecki, and S. Oney, "Edcode: Towards personalized support at scale for remote assistance in cs education," in *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, 2020, pp. 1–5.

[102] Y. Chen, S. W. Lee, and S. Oney, *CoCapture: Effectively Communicating UI Behaviors on Existing Websites by Demonstrating and Remixing*. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/3411764.3445573

[103] R. Krosnick, S. W. Lee, W. S. Laseck, and S. Onev, "Expresso: Building responsive interfaces with keyframes," in *2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, 2018, pp. 39–47.

[104] K. Badillo-Urquiola, Z. Shea, Z. Agha, I. Lediaeva, and P. Wisniewski, "Conducting risky research with teens: Co-designing for the ethical treatment and protection of adolescents," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW3, pp. 1–46, 2021.

[105] N. McDonald, K. Badillo-Urquiola, M. G. Ames, N. Dell, E. Keneski, M. Sleeper, and P. J. Wisniewski, "Privacy and power: Acknowledging the importance of privacy research and design for vulnerable populations," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–8. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/3334480.3375174

[106] A. K. Ghosh, C. E. Hughes, and P. J. Wisniewski, "Circle of trust: A new approach to mobile online safety for families," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–14. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/3313831.3376747

[107] N. Chatlani, Z. Shea, and P. Wisniewski, "Teenovate: Using intergenerational participatory design to teach students about adolescent online safety," in *Chatlani, Neeraj and Shea, Zachary and Wisniewski, Pamela, Teenovate: Using Intergenerational Participatory Design to Teach Students about Adolescent Online Safety (June 18, 2020). Teaching CCI Workshop of the 2020 ACM Interaction Design and Children Conference*, 2020.

[108] Perverted Justice Foundation Inc. (2020) Perverted-justice.com archives. [Online]. Available: http://www.perverted-justice.com/?archive=byUserVotes

[109] P. Zambrano, J. Torres, L. Tello-Oquendo, R. Jácome, M. E. Benalcázar, R. Andrade, and W. Fuertes, "Technical mapping of the grooming anatomy using machine learning paradigms: An information security approach," *IEEE Access*, vol. 7, pp. 142 129–142 146, 2019.

[110] J. D. M.-W. C. Kenton and L. K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.

[111] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018, work in progress.

[112] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

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

[114] 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.

[115] J. Li, M. Galley, C. Brockett, J. Gao, and W. B. Dolan, "A diversity-promoting objective function for neural conversation models," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 110–119.

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

[117] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.

[118] J. Chen and D. Yang, "Multi-view sequence-to-sequence models with conversational structure for abstractive dialogue summarization," *arXiv preprint arXiv:2010.01672*, 2020.

[119] C. Laorden, P. Galán-García, I. Santos, B. Sanz, J. M. Hidalgo, and P. G. Bringas, "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*. Springer, 2013, pp. 261–270.

[120] 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*, vol. 16, pp. 105–133, 2002.

[121] A. Bignold, F. Cruz, R. Dazeley, P. Vamplew, and C. Foale, "An evaluation methodology for interactive reinforcement learning with simulated users," *Biomimetics*, vol. 6, no. 1, p. 13, 2021.

[122] J.-H. Cho, Y. Wang, I.-R. Chen, K. S. Chan, and A. Swami, "A survey on modeling and optimizing multi-objective systems," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1867–1901, 2017.

[123] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, pp. 1–67, 2020.

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

[125] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[126] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *International Conference on Machine Learning*, 2015, pp. 1889–1897.

[127] J. Peters and S. Schaal, "Natural actor-critic," *Neurocomputing*, vol. 71, no. 7-9, pp. 1180–1190, 2008.

[128] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[129] J. R. Wright and K. Leyton-Brown, "Beyond equilibrium: Predicting human behavior in normal-form games," in *Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[130] ——, "Predicting human behavior in unrepeated, simultaneous-move games," *Games and Economic Behavior*, vol. 106, pp. 16–37, 2017.

[131] C. F. Camerer, *Behavioral game theory: Experiments in strategic interaction*. Princeton university press, 2011.

[132] Z. Yu, Z. Xu, A. W. Black, and A. Rudnicky, "Chatbot evaluation and database expansion via crowdsourcing," in *Proceedings of the chatbot workshop of LREC*, vol. 63, 2016, p. 102.

[133] M. Burtsev, V. Logacheva, V. Malykh, I. V. Serban, R. Lowe, S. Prabhumoye, A. W. Black, A. Rudnicky, and Y. Bengio, "The first conversational intelligence challenge," in *The NIPS '17 Competition: Building Intelligent Systems*, S. Escalera and M. Weimer, Eds. Cham: Springer International Publishing, 2018, pp. 25–46.

[134] W. S. Lasecki, R. Wesley, J. Nichols, A. Kulkarni, J. F. Allen, and J. P. Bigham, "Chorus: A crowd-powered conversational assistant," in *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 151–162. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/2501988.2502057

[135] F. Bessho, T. Harada, and Y. Kuniyoshi, "Dialog system using real-time crowdsourcing and twitter large-scale corpus," in *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, ser. SIGDIAL '12. USA: Association for Computational Linguistics, 2012, p. 227–231.

[136] T.-H. K. Huang, J. C. Chang, and J. P. Bigham, *Evorus: A Crowd-Powered Conversational Assistant Built to Automate Itself Over Time*. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–13. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/3173574.3173869

[137] "Participation agreement." [Online]. Available: https://www.mturk.com/participation-agreement

[138] S. Hosio, J. Goncalves, N. van Berkel, S. Klakegg, S. Konomi, and V. Kostakos, "Facilitating collocated crowdsourcing on situated displays," *Human–Computer Interaction*, vol. 33, no. 5-6, pp. 335–371, 2018.

[139] S. W. Lee, Y. Chen, N. Klugman, S. R. Gouravajhala, A. Chen, and W. S. Lasecki, "Exploring coordination models for ad hoc programming teams," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '17. New York, NY, USA: ACM, 2017, pp. 2738–2745. [Online]. Available: http://doi.acm.org/10.1145/3027063.3053268

[140] A. Razi, K. Badillo-Urquiola, and P. J. Wisniewski, "Let's talk about sext: How adolescents seek support and advice about their online sexual experiences," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13.

[141] Z. Agha, N. Chatlani, A. Razi, and P. Wisniewski, "Towards conducting responsible research with teens and parents regarding online risks," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–8.

[142] P. Wisniewski, H. Xu, M. B. Rosson, and J. M. Carroll, "Parents just don't understand: Why teens don't talk to parents about their online risk experiences," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, ser. CSCW '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 523–540. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/2998181.2998236

[143] X. Caddle, A. Alsoubai, A. Razi, S. Kim, S. Ali, G. Stringhini, M. D. Choudhury, and P. Wisniewski, "Instagram data donation: A case for partnering with social media platforms to protect adolescents online," in *ACM Conference on Human Factors in Computing Systems (CHI 2021)/Social Media as a Design and Research Site in HCI: Mapping Out Opportunities and Envisioning Future Uses Workshop*, 2021.

[144] B. K. Nastasi, J. J. Schensul, M. W. A. D. Silva, K. Varjas, K. T. Silva, P. Ratnayake, and S. L. Schensul, "Community-based sexual risk prevention program for sri lankan youth: Influencing sexual-risk decision making," *International Quarterly of Community Health Education*, vol. 18, no. 1, pp. 139–155, 1998. [Online]. Available: https://doi.org/10.2190/D19D-7NHE-8QG9-CC6B

[145] K. Winskell, G. Sabben, V. Akelo, K. Ondeng'e, C. Obong'o, R. Stephenson, D. Warhol, and V. Mudhune, "A smartphone game-based intervention (Tumaini) to prevent HIV among young africans: Pilot randomized controlled trial," *JMIR Mhealth Uhealth*, vol. 6, no. 8, p. e10482, Aug 2018. [Online]. Available: http://mhealth.jmir.org/2018/8/e10482/

[146] P. Wisniewski, H. Jia, H. Xu, M. B. Rosson, and J. M. Carroll, ""preventative" vs. "reactive": How parental mediation influences teens' social media privacy behaviors," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 302–316. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/2675133.2675293

[147] M. J. Muller and S. Kuhn, "Participatory design," *Commun. ACM*, vol. 36, no. 6, p. 24–28, jun 1993. [Online]. Available: https://doi-org.ezproxy.lib.vt.edu/10.1145/153571.255960

[148] E. Estrada, E. Ferrer, and A. Pardo, "Statistics for evaluating pre-post change: Relation between change in the distribution center and change in the individual scores," *Frontiers in psychology*, vol. 9, p. 2696, 2019.

[149] D. Gefen, D. Straub, and M.-C. Boudreau, "Structural equation modeling and regression: Guidelines for research practice," *Communications of the Association for Information Systems*, vol. 4, no. 1, p. 7, 2000.

[150] H. Whittle, C. Hamilton-Giachritsis, A. Beech, and G. Collings, "A review of young people's vulnerabilities to online grooming," *Aggression and Violent Behavior*, vol. 18, no. 1, pp. 135–146, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S135917891200122X

[151] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," Tech. Rep., 2015.

[152] MIT. (2018) The receptiviti API frameworks. [Online]. Available: https://dashboard.receptiviti.com/docs/frameworks-and-measures/#cog-measures

[153] R. van der Goot, "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*. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 201–206.

[154] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, 2020.

[155] D. Vrandečić and M. Krötzsch, "Wikidata: a free collaborative knowledgebase," *Communications of the ACM*, vol. 57, no. 10, pp. 78–85, 2014.

[156] R. Speer, J. Chin, and C. Havasi, "Conceptnet 5.5: An open multilingual graph of general knowledge," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[157] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

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

[159] C. Lin, "ROUGE: A package for automatic evaluation of summaries," in *Text Summarization Branches Out*. Barcelona, Spain: Association for Computational Linguistics, Jul. 2004, pp. 74–81.

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

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

[162] A. Wang, Y. Pruksachatkun, N. Nangia, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "Superglue: A stickier benchmark for general-purpose language understanding systems," *arXiv preprint arXiv:1905.00537*, 2019.

[163] R. Zellers, A. Holtzman, Y. Bisk, A. Farhadi, and Y. Choi, "Hellaswag: Can a machine really finish your sentence?" in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 4791–4800.

[164] S. Welleck, J. Weston, A. Szlam, and K. Cho, "Dialogue natural language inference," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 3731–3741.

[165] T. Cassidy, B. McDowell, N. Chambers, and S. Bethard, "An annotation framework for dense event ordering," Carnegie-Mellon Univ Pittsburgh PA, Tech. Rep., 2014.

[166] Q. Ning, H. Wu, and D. Roth, "A multi-axis annotation scheme for event temporal relations," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 1318–1328.

[167] E. N. Forsythand and C. H. Martell, "Lexical and discourse analysis of online chat dialog," in *International Conference on Semantic Computing (ICSC 2007)*, 2007, pp. 19–26.

[168] C. K. Joshi, F. Mi, and B. Faltings, "Personalization in goal-oriented dialog," *arXiv preprint arXiv:1706.07503*, 2017.

[169] K. Gopalakrishnan, B. Hedayatnia, Q. Chen, A. Gottardi, S. Kwatra, A. Venkatesh, R. Gabriel, and D. Hakkani-Tür, "Topical-Chat: Towards Knowledge-Grounded Open-Domain Conversations," in *Proc. Interspeech 2019*, 2019, pp. 1891–1895. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2019-3079

[170] A. Sordoni, M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J.-Y. Nie, J. Gao, and B. Dolan, "A neural network approach to context-sensitive generation of conversational responses," *arXiv preprint arXiv:1506.06714*, 2015.

[171] Luis von Ahn. (2021) Useful Resources from Luis von Ahn's Research Group. [Online]. Available: https://www.cs.cmu.edu/~biglou/resources/

[172] Intersoft Consulting. (2021) General Data Protection Regulation. [Online]. Available: https://gdpr-info.eu/

[173] K. Badillo-Urquiola, Z. Shea, Z. Agha, I. Lediaeva, and P. Wisniewski, "Conducting risky research with teens: Co-designing for the ethical treatment and protection of adolescents," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. CSCW3, jan 2021. [Online]. Available: https://doi.org/10.1145/3432930

[174] N. McDonald, K. Badillo-Urquiola, M. G. Ames, N. Dell, E. Keneski, M. Sleeper, and P. J. Wisniewski, "Privacy and power: Acknowledging the importance of privacy research and design for vulnerable populations," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–8. [Online]. Available: https://doi.org/10.1145/3334480.3375174

[175] A. M. Walker, Y. Yao, C. Geeng, R. Hoyle, and P. Wisniewski, "Moving beyond 'one size fits all': Research considerations for working with vulnerable populations," *Interactions*, vol. 26, no. 6, p. 34–39, Oct. 2019. [Online]. Available: https://doi.org/10.1145/3358904

[176] Virginia Tech. (2019) Multicultural academic opportunities program. [Online]. Available: https://www.maop.vt.edu/

[177] ——. (2020) Center for the enhancement of engineering diversity (CEED). [Online]. Available: https://eng.vt.edu/ceed/ceed-undergraduate-programs.html

[178] ——. (2020) Mcnair assistantship. [Online]. Available: https://graduateschool.vt.edu/funding/types-of-funding/funding-sponsored-by-the-graduate-school/mcnair-assistantship.html

[179] NCWIT. (2020) The national center for women & information technology (NCWIT). [Online]. Available: https://www.ncwit.org/programs/pacesetter-members/all

[180] Virginia Tech. (2019) Interdisciplinary studies (IDST). [Online]. Available: https://www.undergradcatalog.registrar.vt.edu/0204/ucdCIS.html#id3116397756

[181] X. Zhao, F. Chen, S. Hu, and J. Cho, "Uncertainty aware semi-supervised learning on graph data," in *Advances in Neural Information Processing Systems (NIPS)*, 2020.

[182] X. Zhao, F. Chen, and J. Cho, "Deep learning based scalable inference of uncertain opinions," in *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2018, pp. 807–816.