4.4 Technical Perspectives of Sexual Online Grooming

Jenny Felser[1], Svenja Preuß[1], Dirk Labudde[1] & Michael Spranger[1]

[1]Forensic Science Investigation Lab (FoSIL), Hochschule Mittweida, Mittweida, Germany

With the increasing popularity of social networks and chat rooms, new risks for children and adolescents are emerging. In particular, they are exposed to the danger of sexual online grooming. To address this problem, technical approaches have been developed to assist law enforcement in investigating sexual online grooming and protect young people while chatting. This chapter aims to approach the issue of sexual online grooming from a technical point of view. Research in this area has focused on developing methods to detect sexual online grooming in chats automatically. These methods already achieve a high level of reliability. Thus, they provide essential support for law enforcement agencies in analysing the often immense amount of chatlogs for evidence of sexual online grooming. In addition, technical approaches can assist investigators in undercover work, sometimes required to convict sexual offenders. This includes automatically generating a summary of the linguistic habits of the person whose identity the police officer intends to assume. Moreover, chatbots posing as adolescents to attract sexual offenders have been suggested. Furthermore, this chapter describes how semi-automatic and automatic linguistic methods can be used to analyse the phases of the sexual online grooming process and presents real-time protection tools against sexual online grooming.
Keywords: sexual online grooming, text analysis, automatic detection, protection tools, chatbots

Introduction

Digital communication, such as social networks and chat rooms, is becoming increasingly popular among children and adolescents. However, it also exposes young people to new dangers. In particular, sexual online grooming has developed into a growing risk. In 2021, the German Federal Criminal Police Office reported 3,539 cases of sexual abuse of children using internet communication technologies, which includes sexual online grooming (SOG) (Bundeskriminalamt, 2022). Compared to the previous year, the number has increased by 34.5% (Bundeskriminalamt, 2022).

Therefore, detecting sexual online grooming in chats constitutes an essential task for law enforcement agencies (Ngejane et al., 2021). However, a key challenge is that investigators are often confronted with immense amounts of chat logs to analyse for potential evidence (al-Khateeb &

Epiphaniou 2016). Manually examining the texts is time-consuming and tedious and increases the risk of overlooking essential clues (Anderson et al., 2019). Furthermore, when dealing with a sensitive topic such as SOG, an unintentional bias of the investigator is also a potential source of error (Anderson et al., 2019). In addition, the intensive and continual confrontation with online child sexual abuse can be a psychological burden (Seigfried-Spellar, 2018; Zuo et al., 2018). Consequently, there is a need for tools that can automatically detect SOG in chat logs and, thus, provide support in the investigation process (Ngejane et al., 2021).

The first attempts at automatically detecting SOG were already made in 2007 as part of the research project "Study for the Termination of Online Predators" (STOP) at Iowa State University (Harms & Ferlazzo, 2007; Pender, 2007). A primary goal of this project was to develop a method to identify participants in chats who are potential sexual offenders (Pendar, 2007). Most notably, research in this area was strongly driven by the International Sexual Predator Identification Competition at PAN-2012, where a total of 16 teams competed in various tasks related to identifying SOG in chat logs (Inches & Crestani, 2012).

However, research in the technical field has been broadened beyond the automatic detection of SOG for the assistance of law enforcement agencies. Efforts have also been made to understand the grooming process in more detail, develop real-time protection systems and support investigators' undercover work. This chapter aims to present the different technical approaches that can be used to analyse and detect SOG.

First, we describe how the stages of the SOG process can be examined by employing semi-automatic and automatic linguistic analyses. Then, approaches for the automatic detection of SOG are presented and discussed. Subsequently, this chapter outlines current practical solutions to the problem of SOG. These include automated tools for protecting children during online communication and approaches to assist the undercover work of law enforcement agencies. An overview of the datasets available for research in SOG follows. Finally, we conclude and provide an outlook for future research.

Analysis of the sexual online grooming process using technical approaches

Semi-automatic or automatic linguistic-based text analysis can provide valuable insights into the SOG process. O'Connell (2003) identified five stages of SOG, i.e. friendship-forming, relationship-forming, risk assessment, exclusivity and sexual stage. Using Pennebaker's (2015) Linguistic Inquiry and Word Count (LIWC) tool, which calculates the percentage of words belonging to predefined psychological word categories as defined in the tool's dictionary, Black et al. (2015) and Gupta et al. (2012) detected specific types of characteristics in chat parts. Subsequently, the authors used the results to explore and evaluate O'Connell's (2003) five-stage model of the grooming process. The results suggest that the LIWC tool may be used to detect the different stages in chats.

Black et al. (2015) analysed the language used in these stages based on chat messages written by 44 convicted sexual offenders. Analogous to the five phases, the messages were divided into five equal-sized parts based on word count, which were then analysed with the LIWC tool (Black et al., 2015). For each stage, the authors selected specific LIWC word categories that may be expected to correspond to the respective stage, reflecting the intentions and purposes of the sexual offenders (Black et al., 2015). For instance, the risk assessment stage, in which the sexual offender tries to find out how likely a person from the child's private surroundings will discover him, was represented, for example, by words of the categories "family" and "anxiety" (Black et al., 2015). By examining whether the words in the specific categories of a stage were frequently used in the corresponding part of the chat, the authors found that the vocabulary defined in the LIWC dictionary in some cases does not correspond to the order of the stages defined by O'Connell (2003) (Black et al., 2015).

Gupta et al. (2012) also examined O'Connell's (2003) model using the LIWC tool but in contrast to the previous approach, they manually separated 75 sexual online grooming conversations into individual stages. In addition, the authors applied logistic regression analysis to the LIWC results for each grooming stage, intending to identify LIWC categories that are particularly indicative of a stage (Gupta et al., 2012). For example, they concluded that social category words (e.g., "mate" and "talk") are distinctive of the relationship-forming stage, in which the friendship between the child and the sexual offender becomes more intense (Gupta et al., 2012). The manual segmentation of the chats into stages also allowed the authors to determine the distribution of the stages within the grooming conversations.

They found that the most significant proportion of each conversation is taken up by the relationship-forming stage (Gupta et al., 2012).

Unlike the approaches discussed so far, the approach proposed by Zambrano et al. (2019) is not oriented towards O'Connell's (2003) model. Instead, they developed a model for describing the SOG process, which is inspired by life cycles proposed in the field of information security to describe the phases or steps of a successful computer attack (Zambrano et al., 2019). As a first step, they attempted to identify the stages of SOG using Latent Dirichlet Allocation (LDA) as a topic modelling algorithm and chats from convicted sexual offenders as input (Zambrano et al., 2019). LDA reveals a given number of topics, where the topics are described by a probability distribution over words (Blei, 2003). In the case of Zambrano et al. (2019), six topics were identified. In order to characterise these six topics with linguistic aspects, in the second step, LIWC categories were assigned to a topic if the most probable terms in this topic were included in the corresponding LIWC category (Zambrano et al., 2019). The assigned LIWC categories allowed the authors to derive the sexual offender's intention for each topic or stage, which in turn enabled them to describe the phases by comparing these intentions to those of each stage of selected life cycles of computer attacks mentioned in the literature (Zambrano et al., 2019).

Automatic detection of sexual online grooming

In recent decades, several approaches have been developed to detect SOG automatically. Their aim is primarily to assist law enforcement agencies in the forensic analysis of chat logs and to reduce the time needed to detect past grooming attacks (e.g., Bours & Kulsrud, 2019; Ngejane et al., 2021; Wani et al., 2021). The procedure usually used to achieve this goal is shown in Figure 1. As can be seen, SOG detection consists of several tasks that can be accomplished by employing a hierarchical approach (e.g., Bours & Kulsrud, 2019; Villatoro-Tello et al., 2012).

Usually, the first task is to identify the SOG conversations (see Task 1 in Figure 1) (Villatoro-Tello et al., 2012). Subsequently, a distinction is made between the sexual offender and his victim within the detected predatory conversations (task 2a) (e.g., Borj et al., 2020; Cardei & Rebedea, 2017; Villatoro-Tello et al., 2012). Finally, those messages are determined that are particularly indicative of the grooming process (task 3) (e.g., Peersman

et al., 2012; Tomljanović et al., 2016). In addition, two approaches were proposed that reduce the number of conversations to be studied (tasks 0a and 0b). In the first approach, Siva et al. (2021) suggest using automatic age detection to distinguish minors from adult social media users, primarily based on the assumption that the writing styles differ, for example, in terms of the use of slang and emoticons (task 0a). The second approach (task 0b) consists of examining only the conversations of users who have provided false information about their gender and age in their social media profiles based on the fact that sexual offenders sometimes pose as adolescents and, in some cases, male offenders pretend to be female (Ashcroft et al., 2015; Peersman et al., 2011; van de Loo et al., 2016). In this context, the authors again attempted – corresponding to task 0a – to determine from the chat data whether the users were minors or adults. In order to achieve this, they tried to determine gender based on the users' language style and vocabulary. They then identified the users whose age and gender information in their profiles did not match the predicted age and gender. These two approaches can not only be used for forensic purposes but are also particularly suitable for preventive systems that aim to notify social network moderators of grooming attacks on time (Peersman et al., 2011; Siva et al., 2021; van de Loo et al., 2016). Both approaches in this context aim to reduce the number of conversations that need to be continuously monitored (van de Loo et al., 2016). Furthermore, the first approach (task 0a), where only conversations involving an adolescent are monitored, prevents the intrusion of privacy in conversations between two adults (Siva et al., 2021).

It should be noted that there is no research concentrating on the entire process shown in Figure 1. Instead, several authors focus on one or more tasks (e.g., Pandey et al., 2012; Parapar et al., 2014; Wani et al., 2021; Zuo et al., 2018). For instance, some authors aimed to identify sexual offenders in social media texts without detecting grooming conversations first (task 2b) (e.g., Parapar et al., 2014; Wani et al., 2021).

The individual tasks of grooming detection are usually realized by employing text categorisation, which aims to group text documents into predefined categories (Dalal & Zaveri, 2011). The term "text documents" refers, in this case, to conversations for task 1, all messages a user has written for task 2x or single messages of a user for task 3, and the categories can be, for example, *predatory* and *non-predatory* or *victim* and *sexual predator*. The text documents are assigned to these categories based on their characteristics, known in this context as features.

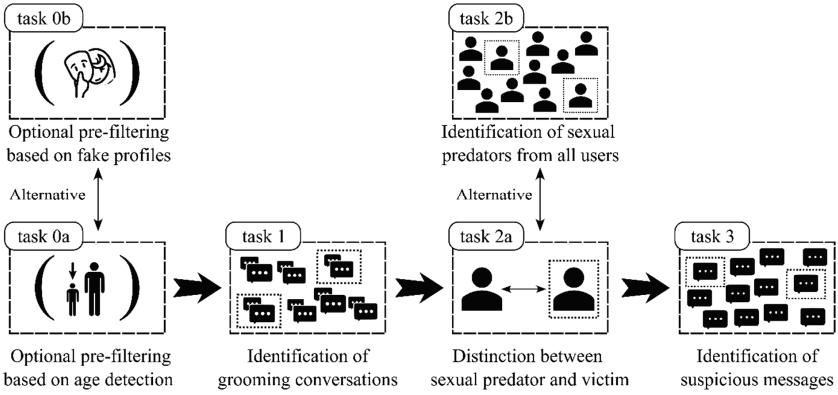

Figure 1: The procedure applied for the automatic detection of sexual online grooming

Notes. No author has applied all steps of the presented procedure.

Features

Previous work has analysed and compared different types of features, which will be described in the following.

Lexical features

Among the most commonly used features are lexical features, which are sometimes also denoted as textual features and capture the content and, respectively, the vocabulary of the conversation texts (Razi et al., 2021). Lexical features were often extracted using the standard Bag of Words (BoW) model (e.g., Anderson et al., 2019; Bours & Kulsrud, 2019; Zuo et al., 2018). BoW models consider a document as a set of its constituent terms, i.e. words, and represent it as a vector of so-called term frequencies (TF) (Salton, 1968). The TF indicates how often a word occurs in the document (Salton, 1968). Furthermore, some works applied the TF-IDF heuristic (e.g., Borj & Bours, 2019; Ngejane et al., 2018; Pendar, 2007), which combines the TF with the inverse document frequency (IDF), which gives more weight to specific words and attempts to capture the importance of a term (Jones, 1972).

Since a well-known problem of BoW is that it does not consider the relationships between words and the word order, phrases were sometimes used

as lexical features instead of individual words (e.g., Borj & Bours, 2019; Pendar, 2007). Besides, some authors included character level n-grams (e.g., Bogdanova et al., 2014; Popescu & Grozea, 2012; Ringenberg et al., 2019), which are understood to be a sequence of consecutive characters (Sapkota et al., 2015). These features have the advantage of being robust against spelling errors and different morphological variants of a word (Mladenović et al., 2021).

In order to address the problem that BoW ignores the semantic and syntactic meaning of words (Shao et al., 2018), it was proposed to use word embedding features. In those features, the words of a document are represented as low-dimensional vectors in such a way that semantically similar words have similar vector representations (Borj et al., 2020). Examples of word embeddings that have been employed to detect grooming are Word2Vec (e.g., Muñoz et al., 2020; Tomljanović et al., 2016) and GloVe vectors (Borj et al., 2020; Ebrahimi et al., 2016).

Finally, some authors created a dictionary containing typical words for grooming (e.g., Kontostathis et al., 2010; Wani et al., 2021). For this purpose, they analysed grooming conversations and identified terms that were frequently used by sexual offenders or by their victims. As lexical features, for example, the frequency of these dictionary words in the text documents was chosen (Wani et al., 2021).

Behavioural and stylistic features

In addition, behavioural features that describe characteristics of sexual offenders' chats and a participant's actions in conversations were used primarily for tasks 2a and 2b (see Figure 1) (Cardei & Rebedea, 2017). These features include a participant's response time (Morris & Hirst, 2012), the usual time of day a participant chats, the number of individuals contacted (Parapar et al., 2014) and the number of conversations started by a participant (Dhouioui & Akaichi, 2016).

Since a person's writing style also characterises their behaviour, the boundaries between behavioural and stylistic features are blurred. Features that are more stylistic in nature range from the proportion of slang words used by a conversation participant (Cardei & Rebedea, 2017) to the number of emoticons (Dhouioui & Akaichi, 2016) and imperative sentences (Bogdanova et al., 2014) to the average word length (Pandey et al., 2012) in a user's messages or conversation.

Moreover, Cardei and Rebedea (2017) introduced features that reflect the interaction between participants in a conversation and capture differences in their behaviour and writing style. For instance, they compared the proportion of questions, negations and misspelt words in a user's messages with those of their interlocutors.

Syntactical features

Another important aspect of a person's writing style is the usage of parts of speech (POS), which can be described by syntactical features (Cano et al., 2014). Here, either all detected POS within a message (Cano et al., 2014) or the (relative or weighted) frequency of certain word types (Bogdanova et al., 2014; Pandey et al., 2012) were used as features. In this context, Bogdanova et al. (2014) highlighted that personal pronouns, reflexive pronouns and modal verbs of obligation, such as "have to", "must", and "shouldn't", are particularly appropriate syntactical features for identifying predatory conversations (task 1, Figure 1) (Bogdanova et al., 2014). The increased use of these POS may indicate neuroticism (Argamon et al., 2009), which is often more prevalent in sexual offenders (Carvalho & Nobre, 2019).

Sentiment features

Based on the assumption that sexual offenders are generally considered emotionally unstable and suffer from mental health problems (Briggs et al., 2011; Nijman et al., 2009), some authors tried to reveal the emotional state of participants of a conversation using sentiment features (Bogdanova et al., 2014; Cano et al., 2014; Cheong et al., 2015; Wani et al., 2021). These features were determined using dictionaries with words associated with a positive or negative sentiment or a particular emotion, such as sadness, joy and anger. As features, for instance, the number of words belonging to a particular emotion category were used (Bogdanova et al., 2014; Wani et al., 2021).

LIWC features

Finally, psycho-linguistic patterns were considered features (Cano et al., 2014; Parapar et al., 2014) that can be identified with the LIWC tool presented in the previous section. Features based on the LIWC tool aim to capture a communicator's matters and interests as well as their psychological state and affective characteristics (Parapar et al., 2014). Examples of categories so far are the categories of words related to money (e.g., "cash", "owe"), to family (e.g., "daughter", "aunt") and to sex (e.g., "horny", "incest") and various emotional categories, including words that express anxiety and sadness (Cano et al., 2014; Parapar et al., 2014). Further categories refer to specific POS, such as pronouns, prepositions and auxiliary verbs (Parapar et al., 2014). Thus, overlaps with the syntactical and sentiment features exist.

Methods

These features are particularly needed for text classification by supervised machine learning, which is the most used approach for detecting SOG. In addition, rule-based classification systems and unsupervised machine learning methods, namely clustering, have been applied so far. The following subsection describes these three approaches in detail, which are also illustrated in Figure 2, where they were assigned to the tasks shown in Figure 1, for which they were primarily used.

Supervised machine learning algorithms

Supervised machine learning models can learn by example to classify the text documents into the categories mentioned at the beginning of this section, such as *predatory* and *non-predatory*, based on discriminative features (Zhai & Massung, 2016). For this, the models must be trained on large labelled datasets consisting of text documents and their corresponding categories (Zhai & Massung, 2016). After the training phase, they are able to make a prediction about the category of new, unlabelled text documents (Zhai & Massung, 2016). In SOG detection, mainly traditional machine learning algorithms, such as Support Vector Machines (e.g., Pandey et al., 2012; Parapar et al., 2014; Tomljanović et al., 2016), k-Nearest Neighbour (e.g., Pendar, 2007), Naïve Bayes (e.g., Bours & Kulsrud, 2019), Random

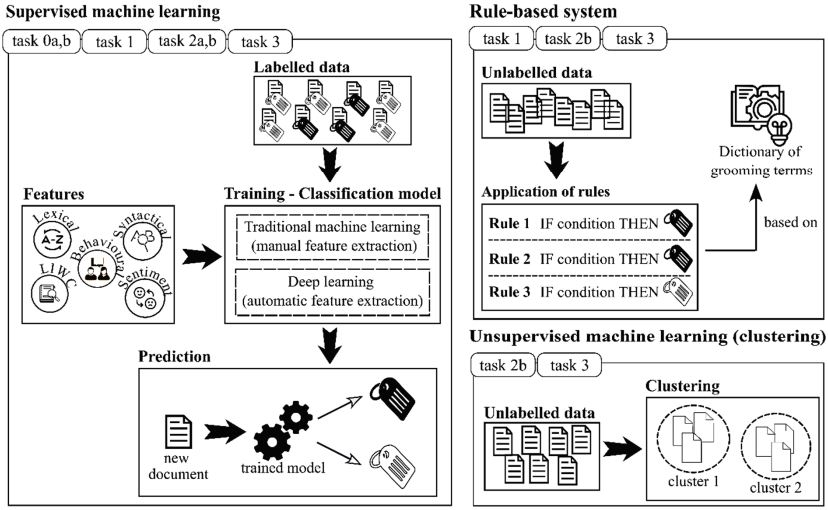

Figure 2: Methods used for the individual tasks of automatic sexual online grooming detection

Forest (e.g., Zuo et al., 2018) and Logistic Regression (e.g., Anderson et al., 2019), have been used.

However, deep learning models based on deep stacked artificial neural networks, which mimic processes of the human nervous system, are also gaining popularity (Yasaka et al., 2018). Convolutional Neural Networks (CNN) (e.g., Ebrahimi et al., 2016; Misra et al., 2019; Muñoz et al., 2020), Recurrent Neural Networks (RNN) (Kim et al., 2020) and Long Short-Term Memory (LSTM) (Ngejane et al., 2021) have been applied by now. Deep learning models differ from traditional machine learning models in that they can perform so-called automatic feature extraction. Based on this, Preuß et al. (2021) proposed an approach to address tasks 1, 2a and 3 of grooming detection, shown in Figure 1. The basic idea was to use a CNN to automatically extract the most important lexical features from the conversation texts. As reported before, previous approaches have tended to define lexical features based on a hand-crafted dictionary of terms characteristic of the grooming process. However, one challenge is defining which words are relevant for detecting predatory chats. This problem can be approached by CNN, which can learn lexical features appropriate for distinguishing predatory from non-predatory texts. These lexical features are, in this case, a specified number of semantically similar terms (Jacovi

et al., 2018), which are denoted as n-grams (Preuß et al., 2021). A major benefit of this approach is that, for example, regarding task 1, the CNN does not only identify n-grams that frequently occur in predatory conversations (Preuß et al., 2021). Instead, it also learns n-grams that are often used in everyday chats but rarely in the conversations between sexual offenders and their victims and are consequently suitable for distinguishing between suspicious and non-suspicious conversations. These automatically extracted features were used as input for training a further neural network, a multilayer perceptron, which serves as the actual classifier. Furthermore, for all tasks, except identifying the suspicious messages (task 3), behavioural and sentiment features were taken into account in addition to the lexical features.

Rule-based manual approaches for classification

The main drawback of supervised machine learning algorithms, especially deep learning methods, is that an immense amount of labelled data is required (Tyagi & Rekha, 2019). However, publicly available annotated datasets for detecting SOG, especially at the message level (task 3), are limited, and the manual labelling of datasets is time-consuming (Mladenović et al., 2021). An alternative to supervised machine learning algorithms for text classification, which does not require labelled datasets, is to manually create rules that contain conditions under which a text document is assigned to a specific category (Zhai & Massung, 2016). These rules are based on dictionaries consisting of terms and phrases commonly used by sexual offenders in SOG conversations (Gunawan et al., 2016; McGhee et al., 2011; Vartapetiance & Gillam, 2012). For instance, Vartapetiance and Gilliam (2012) defined the following four categories of keywords for the identification of sexual offenders (task 2b):

- phrases used to ask for the minor's address (e.g., "the address"),
- phrases to refer to the child's parents (e.g., "your mom"),
- phrases to emphasise the age difference between the sexual offender and the child (e.g., "you are young") and
- phrases to express the desire for a sexual approach (e.g., "go down on you").

Accordingly, a user was classified as a sexual offender if he utilised the phrases of these categories in a predefined occurrence.

Unsupervised machine learning algorithms (clustering)

Finally, besides the supervised machine learning algorithms mentioned above, few works have used clustering (Kodžoman et al., 2016; Toriumi et al., 2015) as a typical unsupervised machine learning method that does not require labelled data (Alloghani et al., 2020). The task of clustering is to detect groups of similar objects in a dataset (Alloghani et al., 2020).

For instance, Kodžoman et al. (2016) clustered all messages written by a user who had previously been classified as a sexual offender to identify those lines that were particularly indicative of his bad behaviour (task 3). Therefore, they applied the k-means clustering algorithm, described by Hartigan and Wong (1979), to detect two clusters of similar messages (Kodžoman et al., 2016). Subsequently, they determined which clusters contained the suspicious messages and which were the irrelevant ones.

Moreover, Toriumi et al. (2015) addressed task 2b, the identification of sexual offenders, by clustering. For this purpose, they used the Gaussian Mixture Model, introduced by Bishop (2006), to cluster users of private chat systems based on their communication behaviour (Toriumi et al., 2015). Private chat systems give users who have first met in multiparticipant chat systems the opportunity to exchange messages secretly in one-to-one chats and are, therefore, particularly attractive for sexual offenders to lure potential victims (Toriumi et al., 2015). The authors identified clusters of users that could be either sexual offenders or grooming victims, both characterised by highly active communication behaviour (Toriumi et al., 2015). They assumed that sexual offenders are users who start one-to-one chats with many other users, whereas those users who are contacted by many other users and receive many messages have a higher risk of becoming grooming victims (Toriumi et al., 2015).

Results and discussion

The final subsection compares different approaches to detecting SOG and discusses their results. Therefore, it is first necessary to understand how the performance of a method can be assessed.

Evaluation of approaches for the automatic detection of sexual online grooming

Those researchers who applied clustering methods tended to assess and interpret the results subjectively (e.g., Toriumi et al., 2015). In contrast, the performance of classification systems, i.e. supervised machine learning methods and rule-based systems, was measured quantitatively by calculating evaluation metrics, including precision, recall and F_β-score on a labelled test set (e.g., Cardei & Rebedea, 2017; Gunawan et al., 2016). For this purpose, the classification models' predictions about the documents' categories in the test set are compared with their correct labels. "Precision" describes the probability that a text document that is classified as the category of interest, in this case mostly *predatory*, is actually predatory. In contrast, "recall" indicates how many documents in the class *predatory* were detected by the classification algorithm. The F_β-score combines precision and recall using the weighted harmonic mean, where the parameter β controls whether precision or recall is deemed more important for the given task. As a standard, this parameter is set to one so that recall and precision are equally weighted. However, especially in the case of sexual offender identification (task 2a and task 2b), several researchers have used the $F_{0,5}$-score, which emphasises precision (Cardei & Rebedea, 2017; Fauzi & Bours, 2020; Wani et al., 2021). Inches and Crestani (2012) justified this with the fact that it is more important that law enforcement agencies are presented with the right suspects than with all possible ones in order to reduce their required time to identify sexual offenders. Regarding task 3, the F_3-score was often preferred (e.g., Kodžoman et al., 2016; Preuß et al., 2021; Tomljanović et al., 2016), which gives higher weight to recall (Borj et al., 2020). Accordingly, the focus was on recognising as many relevant lines as possible to gather the maximum amount of evidence against a suspect (Inches & Crestani, 2012).

Overview of the performance of the previous approaches

The fact that the authors used different metrics and different data sets to evaluate the performance of their methods makes it difficult to assess and compare the individual approaches. Nevertheless, it can be concluded from the results that task 1 and task 2a or task 2b of automatic SOG detection can be performed automatically with high reliability. A performance of up to 99% in the $F_{0.5}$-score can be achieved in detecting SOG conversations

(task 1) (Fauzi & Bours, 2020). Current approaches can also obtain high results in the subsequent identification of the sexual offenders in these conversations (task 2a), with $F_{0.5}$-scores of up to 93% (Fauzi & Bours, 2020). Furthermore, an $F_{0.5}$-score of 96% was achieved in identifying sexual offenders from all users (task 2b) (Wani et al., 2021). The performance for task 3, identifying the most distinctive lines for grooming behaviour, is lower with F_3-scores of up to 67% (Preuß et al., 2021). The approaches from which the performance measure was reported all used the PAN-2012 dataset for training and evaluation of their models (Inches & Crestani, 2012), which is described in the penultimate section.

Regarding the optional pre-filtering (task 0), it should be noted that the results for the distinction between underaged and adults, which is required for both subtasks (task 0a and task 0b), depend strongly on the chosen age limit at which a person was considered as adult. The authors set this age limit differently (Peersman et al., 2011; Siva et al., 2021; van de Loo et al., 2016). Nevertheless, the F_1-score of 84% obtained by Siva et al. (2021) in classifying users as under or over 18 on a dataset of conversations from different adult and children chat systems indicates that these approaches have the potential to pre-select conversations for subsequent grooming classification. Regarding performance in the automatic detection of gender, van de Loo et al. (2016) obtained an F_1-score of 75% on a dataset consisting of chat posts from the Belgian social network Netlog when *female* was the category of interest and an F_1-score of 58% with *male* as the category of interest.

Comparison of the used features

Concerning the different types of features used, several studies have found that simple lexical features are suitable to differentiate between the sexual offender and his victim in a predatory conversation (task 2a) (e.g., Bours & Kulsrud, 2019; Fauzi & Bours, 2020; Pendar, 2007). Pendar (2007) concluded that sexual offenders have a characteristic vocabulary that distinguishes them from their interlocutors. In contrast, lexical features were considered insufficient by Pandey et al. (2012) and by Bogdanova et al. (2014) to recognise SOG conversations from legal chats on sexual topics between adults. Bogdanova et al. (2014) suggested that, in the more complicated cases of discriminating "cyber pedophiles's" conversations from cybersex chat logs, a combination of different features, such as sentiment features,

psycho-linguistic features and syntactical features, should be used instead. Several studies have found that improvements over a single feature set can be achieved by combining different feature types. For example, addressing task 2b, Wani et al. (2021) combined sentiment features and lexical features and Parapar et al. (2014) lexical features, behavioural features and psycho-linguistic features. In particular, the analysis of behavioural features, for example, using the statistical distribution of features in the sexual offenders' and victims' chats (Morris & Hirst, 2012), also offers the opportunity to gain insights into the characteristics and behavioural patterns of sexual offenders. For example, studies considering behavioural features revealed that sexual offenders tend to dominate a conversation and are characterised by high user activity, starting conversations frequently, sending many messages (Morris & Hirst, 2012) and responding after a short time (Parapar et al., 2014).

Comparison of the methods

Considering the results of the different methods investigated, it can be said that, in accordance with the so-called No Free Lunch Theorem (Wolpert & Macready, 1997), no single algorithm consistently produces the best result. However, different algorithms have to be considered appropriate, depending on the data set, the selected features and the task. Both traditional machine learning algorithms, such as the Support Vector Machine (e.g., Borj et al., 2020) and Random Forest (e.g., Cardei & Rebedea, 2017), and deep learning methods (Kim et al., 2020) yielded promising results.

To our knowledge, rule-based systems cannot outperform machine learning approaches when evaluated on the same dataset (PAN-2012 dataset). Their performance was less than 55% in terms of $F_{0.5}$-score in identifying sexual offenders (task 2b) (e.g., Vartapetiance & Gillam, 2012; Vilariño et al., 2012) and 42% or less in terms of F_3-score in marking the most relevant lines for the grooming process (task 3) (e.g., Kontostathis et al., 2012). One of the major limitations of rule-based systems pointed out by Kontostathis et al. (2012) is that they fail to detect subtle, less explicit sexual innuendoes. In addition, rule-based systems require that categories are clearly defined (Zhai & Massung, 2016). However, this is not always the case in SOG detection because SOG conversations may contain words also found in conversations of sexual nature between adults and common chats about arranging meetings (Kontostathis et al., 2012).

Moreover, rule-based systems are inflexible and rigid, so even if they have shown promising results for a particular data set, new data may require adaptation of the rules or the creation of new rules (Zhai & Massung, 2016). For example, Hidalgo and Díaz (2012) attempted to apply an existing rule-based system for detecting suspicious messages in Spanish (task 3) to classify English messages mainly through automatic translation, resulting in an F_3-score of 0%. This result was attributed, among other things, to the lack of rules taking into account typical phrases for English grooming conversations and to the fact that the system was designed for grooming cases where the sexual offender took months to lure his victim slowly. In contrast, in the English dataset, sexual intentions often became apparent after a short time (Hidalgo & Díaz, 2012). This example thus underlines that rule-based systems are not easily transferable to slightly different application scenarios.

Regarding clustering, it is impossible to draw a meaningful comparison between this method and the other two approaches for SOG detection (see Figure 2) since only very few approaches were based on it (Kodžoman et al., 2016; Toriumi et al., 2015). The main disadvantage of clustering is that the researcher must interpret which of the created clusters, containing, e.g., messages, are considered suspicious (Meyers, 2000). In contrast, in the case of classification, they are provided with direct information about whether a message belongs to the relevant or irrelevant category. Apart from an additional error-proneness due to the subjective interpretation, clustering is unsuitable for a fully automated pipeline for grooming detection.

Early detection of sexual online grooming

The approaches presented so far have investigated the problem of automatic grooming detection from a forensic perspective and have focused on classifying complete conversation sequences as *grooming* or *non-grooming* (Milon-Flores & Cordeiro, 2022). Consequently, they are only suitable for detecting grooming attacks that have already been performed (Milon-Flores & Cordeiro, 2022). However, from the point of child protection, it is of higher importance to develop methods that address the detection of SOG in a preventive way (Milon-Flores & Cordeiro, 2022). It is, therefore, necessary to detect SOG attempts as early as possible in order to trigger an alert while the communication between the sexual offender and his victim is still ongoing and especially before an actual physical encounter has been arranged (Milon-Flores & Cordeiro, 2022).

Few methods have been developed for this scenario, known as early text classification (e.g., Escalante et al., 2016, 2017; López-Monroy et al., 2018; Milon-Flores & Cordeiro, 2022). Several authors focused on finding a more appropriate document representation than standard approaches such as BoW to model the short partial conversations that contain only minimal information (Escalante et al., 2017; López-Monroy et al., 2018; Milon-Flores & Cordeiro, 2022). Furthermore, in previous work, Escalante et al. (2016) addressed the problem of early text classification by adapting Naïve Bayes, a traditional supervised machine learning algorithm, for this scenario. Besides, Milon-Flores and Cordeiro (2022) identified several sentiment, behavioural and stylistic features, including the start time of a conversation and the proportion of correctly spelt words, which they considered particularly suitable for the early detection of SOG. López-Monroy et al. (2018) demonstrated that it is already possible to reliably identify a conversation with an F_1-score of 94% as SOG when only 50% of the course of the conversation is known.

Solutions in practice

Solutions in practice to the problem of SOG include both automated tools designed to increase the safety of children while chatting online and approaches to support the work of law enforcement agencies.

Tools for the protection against sexual online grooming

First, software tools and frameworks to protect children from sexual offenders are presented. While the focus has tended to be on the development of general parental control software, such as Bark (Bark, 2022), FamiSafe (Wondershare, 2022) and Qustodio (Qustodio LLC, 2022), which allow parents to monitor their child's communication on social media platforms, emails or list of contacts (Winters & Jeglic, 2022), there are only few software tools that focus specifically on the protection against SOG. Moreover, these are mostly limited to the English language.

One such system is SafeChat, developed by MacFarlane and Holmes (2016), an automatic monitoring and security system that can be integrated into any instant messaging and communication service. The system checks every message the child intends to send to another user if it contains

personal information, such as their address or phone number, and blocks it. In addition, messages in which the child arranges to meet with another user are blocked. The detection of appointments is based on a comprehensive ontology, which is generally defined as a model for representing knowledge in a domain (Gruber, 1995). In this case, it captures the vocabulary commonly used for meeting arrangements in a conversation and describes the relationships between the terms in that vocabulary (MacFarlane & Holmes, 2016). In addition to blocking the messages, a notification is sent to the parents and a warning message to the child (MacFarlane & Holmes, 2016). The authors' studies demonstrated that their system is already successful in preventing the transmission of personal data (MacFarlane & Holmes, 2016) while they continue to work on improving the reliability of meeting arrangement detection (MacFarlane & Holmes, 2016).

Another framework for early detection of grooming, the Grooming Attack Recognition System (GARS), was developed by Michalopoulos et al. (2014). This system continuously monitors a child's internet communication and calculates a score that describes the current risk of grooming to which the child is exposed. After each new message, this score is updated. As soon as the computed risk value exceeds a threshold that is determined based on the age and gender of the child (InfoSec, 2022), a warning message is sent to the parents, and the child is notified of the current danger via a coloured signal (Michalopoulos et al., 2014). For the determination of the risk score, the results of different methods are combined, including classification by supervised machine learning, as described in the previous section, and the automatic identification of the child's personality type and interlocutor (Michalopoulos et al., 2014). In addition, the development of the risk score in the course of the conversation with the respective interlocutor is also taken into account, as well as the amount of time the child's user profile was visible online (Michalopoulos et al., 2014). The authors' experimental studies indicated that one issue with the system is the high number of false negatives, which means that actual grooming attacks do not raise an alarm, yet the results are also highly dependent on the appropriate threshold of alarm activation (Michalopoulos et al., 2014).

Furthermore, Penna et al. (2010, 2013) aimed to protect adolescents from the increasing risk of SOG in Massive Multiplayer Online Games (MMOG), which are usually understood as internet games in which hundreds or thousands of participants play against each other or together against the game (Barnett & Coulson, 2010; Yahyavi & Kemme, 2013). For this purpose, they developed a prototype system installed on a child's

computer that automatically detects when the child arranges an actual physical meeting in the online game chat (Penna et al., 2013). In order to do so, the tool records all conversations during the online game (Penna et al., 2013). Subsequently, each stored message is checked for indicators of an appointment, where indicators comprise words and phrases such as "meet you at" as well as regular expressions. With the help of regular expressions, stereotypical sequences, such as addresses, can be described. Based on the presence of these indicators, it is determined whether a message is suspicious while additionally taking into account whether it contains indicators that argue against SOG, for example, words and phrases that are typical for the game or general English (Penna et al., 2013). In the case of a suspicious message, the child's parents are notified by email, for example (Penna et al., 2013). The authors' experiments, which consisted of inserting simulated conversations about arranging a meeting into the MMOG "World of WarCraft" game chats, yielded that in 88% of these scenarios, the corresponding messages triggered an alarm (Penna et al., 2013).

Finally, AiBA, a tool for the continuous monitoring of chats and the real-time detection of SOG, has been developed based on research by the Norwegian University of Science and Technology (AiBA AS, 2022). AiBA targets both social networking platforms, where the tool is designed to support moderators in detecting sexual offenders in their chat rooms, and private persons, where AiBA is installed on their devices and then monitors their chats (NTNU, 2022). On the one hand, the system automatically detects fake profiles (AiBA AS, 2022), as described above. For this purpose, the age and gender of a user are predicted based on their writing style and typing rhythm (Raffel et al., 2020). As a result, AiBA can reveal when an adult pretends to be a child or their predicted gender does not match the stated gender in their profile (Raffel et al., 2020).

On the other hand, AiBA also applies classification methods for early detection (AiBA AS, 2022). Therefore, the tool determines a risk score for each message and informs the user on a dashboard that a current conversation is likely to be SOG as soon as the aggregated risk value of the previous messages of the conversation exceeds a threshold value (AiBA AS, 2022). A unique feature is that AiBA considers not only text messages to determine the risk value but also voice messages and images (AiBA AS, 2022). On average, only 20 messages of the conversation are sufficient for the identification of SOG (AiBA AS, 2022).

Support of law enforcement agencies

Furthermore, efforts have been made to support law enforcement agencies in undercover online policing, which may be required to identify and apprehend sexual offenders (MacLeod & Grant, 2017). One possible scenario is that police officers have to assume a child's or adolescent's identity after it has been revealed that this person has been the victim of SOG (Grant & MacLeod, 2016). The police officers then impersonate the child to maintain an online conversation with the sexual offender, gather further evidence against him and arrest him (Grant & MacLeod, 2016). Beyond that, it may also be necessary to take over the identity of an arrested sexual offender to investigate his contact network and arrest other sexual offenders (Grant & MacLeod, 2016).

In this context, the UK-based project "Assuming Identities Online" was established under the leadership of Aston University (MacLeod & Grant, 2016). The focus of this project is on describing the linguistic persona of an individual in such a way that it can be assumed by an Undercover Officer (UCO) (MacLeod & Wright, 2020). In order to reduce the preparation time usually needed before adopting an online identity, the software tool "IDentik" was developed as part of this project (MacLeod & Grant, 2017; MacLeod & Wright, 2020). This tool automatically creates a linguistic summary about an individual based on their conversation record, which includes, for example, information about their habits regarding capitalisation, punctuation, variant spellings of words and patterns of turn-taking (MacLeod & Grant, 2017; MacLeod & Wright, 2020). In addition, the tool provides a "translation" of the UCO's language into the chosen individual's language (MacLeod & Wright, 2020). However, MacLeod and Grant (2017) recommended not relying entirely on the software's results, as the UCO must also be able to recognise important aspects of the individual's language use from their chat logs on their own. Therefore, another important component of the project was to enhance the UK's national "Pilgrim" training program for UCO, as described in (HMIC, 2014), with linguistic elements (MacLeod & Grant, 2017).

Chatbots

Another approach that can be used to convict sexual offenders is that police officers or volunteers pose as adolescents in chat rooms from the beginning

and serve as decoys for the sexual offenders (Callejas-Rodríguez et al., 2016). One of the best-known organisations that followed this approach is the American Perverted Justice Foundation, which will be described in more detail in the next section (Perverted Justice Foundation, 2022).

However, one problem with this approach, as highlighted by Callejas-Rodríguez et al. (2016), is that police officers and volunteers will never be sufficient to discover all sexual offenders. In addition, pretending to be a pseudo-victim over a more extended time can be emotionally stressful (Callejas-Rodríguez et al., 2016). Therefore, chatbots – i.e., machine dialogue systems that can communicate with humans in natural language (Callejas-Rodríguez et al., 2016) – were proposed to assist law enforcement agencies (Callejas-Rodríguez et al., 2016; Laorden et al., 2012; Sunde & Sunde, 2021). In countering SOG, chatbots, police officers and volunteers pose as children or adolescents in social networks and chat rooms (e.g., Callejas-Rodríguez et al., 2016; Laorden et al., 2012). For instance, Callejas-Rodríguez et al. (2016) developed a chatbot that mimics the language of adolescents in Mexican Spanish. This chatbot works based on conversational rules that specify which words and phrases in the interlocutor's messages trigger which response from the chatbot. The experimental results by the authors revealed that the chatbot could generate messages similar in the diversity of vocabulary and syntax to those a human adolescent would formulate. The main limitation of their system, however, is the lack of a mechanism to automatically detect whether the chatbot is currently talking to a sexual offender.

This issue is approached by the chatbot developed by Laorden et al. (2012) called Negobot, also known as Lolita (BBC, 2013). Negobot pretends to be an adolescent, more precisely a 14-year-old girl (BBC, 2013) and automatically assesses whether the current conversation partner has sexual intentions (Laorden et al., 2012). As soon as the bot has concluded that the user it is talking to is a potential sexual offender, it alerts the responsible authorities and sends them the entire recorded conversation with the corresponding person (Laorden et al., 2012). The particularity of the system can be seen in the fact that it applies methods of game theory by regarding the conversation as a competitive game (Laorden et al., 2012). Similar to a game player, the chatbot attempts to find the best (conversation) strategy depending on the behaviour of its opponent, i.e. its communication partner, to achieve its goal of gathering enough information to identify them as a sexual offender without arousing suspicion (Laorden et al., 2012). In contrast to the previously presented system, Negobot is language-indepen-

dent (Laorden et al., 2012). So far, the chatbot has been tested on Google's chat service (Kent, 2013), where it was found that Negobot already has extensive conversational capabilities but that improvements are still needed, for example, concerning the recognition of irony (Kent, 2013) and topic changes (Laorden et al., 2012).

However, the legal framework conditions of the respective country must be considered when using such chatbots.

Datasets

As mentioned above, sufficiently large data sets are required for training models to detect SOG automatically. Almost all previous work was based on two data sets in English: chats from the Perverted Justice website (Perverted Justice Foundation, 2022) and the PAN-2012 dataset (Inches & Crestani, 2012), which is, however, based on the Perverted Justice data.

Perverted Justice data

The Perverted Justice Foundation Inc., usually referred to as Perverted Justice or abbreviated as PeeJ, is an American non-profit organisation that, from its inception in 2003 until early 2019, trained volunteers to pose as minors in chat rooms in order to serve as decoys for sexual offenders (Faraz et al., 2022; Perverted Justice Foundation, 2022). Close cooperation with law enforcement agencies was intended to enable the arrest and conviction of sexual offenders (Perverted Justice Foundation, 2022). For this purpose, the organisation provided law enforcement agencies with recorded conversations as evidence. In addition, meetings between the sexual offender and the pseudo-victim offered an opportunity for arrest (Perverted Justice Foundation, 2022). In the case of a conviction of the sexual offender, the conversation between him and the pseudo-victim was published on the website of Perverted Justice (Perverted Justice Foundation, 2022). A total of 622 chat logs of convicted sexual offenders are publicly available (Perverted Justice Foundation, 2022).

Although many researchers have used the Perverted Justice data (e.g., Bogdanova et al., 2014; Gunawan et al., 2016; Pandey et al., 2012; Pendar, 2007), its suitability as a data basis for the creation of systems for automatic grooming detection is controversial. On the one hand, a major criticism is

that the conversations were not conducted with actual victims (e.g., Cheong et al., 2015; Kontostathis et al., 2010). This can be particularly problematic for distinguishing between a victim and a sexual offender (see Figure 1, task 2a) because, as pointed out by Ashcroft et al. (2015), the writing style of actual minor victims and pseudo-victims posing as children can differ. On the other hand, it is at least ensured that the conversations are SOG with an actual sexual offender involved, as the chats were only published after the conviction (Kontostathis et al., 2010). Moreover, it is difficult to gain access to chats with real victims, as law enforcement agencies are often unwilling to hand them over or, depending on the legal framework of the respective country, are not allowed to do so (Inches & Crestani, 2012).

If the data was used to train a model for the distinction between *grooming* and *non-grooming* conversations (see Figure 1, task 1), the Perverted Justice chats were taken as examples of suspicious conversations. In contrast, either chats about common topics (Kontostathis et al., 2010) or legal chats about sex, for example, cybersex conversations between adults (Bogdanova et al., 2014) or posts on pornographic sites (e.g., Gunawan et al., 2016), were used as *non-grooming* conversations.

PAN-2012 dataset

The second major data source is the dataset provided by the organisers of the International Predator Identification competition as part of PAN-2012 (Inches & Crestani, 2012), a series of scientific events and contests, among other things, in digital forensic text analysis (Webis Group, 2022). This dataset consists of Perverted Justice data as *predatory* chat texts and two types of *non-predatory* chats: Internet Relay Chat (IRC) logs and Omegle chat logs (Inches & Crestani, 2012). The chat logs of IRC, which allow users to communicate in so-called channels about specific topics via short text messages (Ziegler, 2004), are not sexual and comprise general discussions (Inches & Crestani, 2012). In contrast, conversations on the website Omegle, which randomly connects two adult users in a one-to-one chat, are predominantly about sexual matters and, accordingly, more challenging to distinguish from SOG (Inches & Crestani, 2012; Ngejane et al., 2018). The PAN-2012 dataset contains significantly more *non-grooming* than *grooming* conversations, accounting for less than 4% (Inches & Crestani, 2012), in order to be as close to reality as possible.

Conclusion

Sexual online grooming has become an increasing problem and poses a particular risk to children and adolescents. To counter this issue, technical approaches can offer valuable support in obtaining insights into the process of SOG, assist criminal prosecution and increase protection against SOG.

Research in this area has mainly focused on assisting law enforcement agencies in reducing the time required to analyse the vast number of chat logs from social media. Therefore, methods have been developed that automatically identify SOG conversations, suspicious users, i.e. potential sexual offenders, or the most relevant messages for the grooming process. For these purposes, mainly supervised machine learning algorithms have been used, which have already achieved high performance in experimental studies, especially in detecting grooming conversations and identifying sexual offenders. The selection of suitable features has a considerable influence on the results, whereby combinations of different types of features, including lexical features, behavioural features and psycho-linguistic features, have proven particularly promising.

In addition, the safety of children while chatting can be improved by systems that monitor a child's chats and alert the parents as soon as possible if SOG is suspected. However, compared to a high number of general parental control systems, few tools have been developed so far that focus specifically on protection against grooming. Unfortunately, these predominantly provide support only for the English language. They either also use supervised machine learning techniques or are based on checking whether phrases indicating SOG can be found in the chat messages.

Another field where technical approaches can assist in countering SOG is undercover policing. Here, tools can automatically provide a police officer with information about the language use of a person whose identity they need to assume. Moreover, chatbots have been developed that mimic the language of children and adolescents to serve as bait for sexual offenders. However, further improvements to the chatbots' communication capabilities and clarification of the legal conditions for use in the respective country are required before actual practical use.

The main challenge at present is the limitation of appropriate data sets that are needed both for developing systems to detect grooming and, for example, for studies on the stages of the grooming process. Almost all existing studies were based on conversations in English provided by the organisation Perverted Justice, where the sexual offender communicated with a

volunteer posing as a child. A possible approach to generate additional data sets in other languages could be to consider the recorded conversations of the presented chatbots. However, conversations with actual victims would be of higher interest but are hardly available, mainly for legal reasons.

Furthermore, almost all approaches to detecting SOG for law enforcement and prevention purposes only consider text messages. An interesting approach for future research, which has so far only been used by one tool for detecting SOG, is the inclusion of voice messages and images, which are often necessary to understand the context of a text message in a conversation. In addition, as sexual offenders sometimes request revealing or nude photos in their conversations with the victim, a combination with existing procedures for automatically detecting child pornography could prove helpful.

References

AiBA AS (2022). *Real-time detection of fake profiles, grooming and toxicity*. AiBA – Safe Digital Lives. https://aiba.ai/ (retrieved December 2, 2022).

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In M. W. Berry, A. Mohamed & B. W. Yap (Eds.), *Supervised and Unsupervised Learning for Data Science* (pp. 3–21). Springer Nature. https://link.springer.com/content/pdf/10.1007/978-3-030-22475-2.pdf

Anderson, P., Zuo, Z., Yang, L., & Qu, Y. (2019). An Intelligent Online Grooming Detection System Using AI Technologies. *Proceedings of the 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–6. https://doi.org/10.1109/FUZZ-IEEE.2019.8858973

Argamon, S., Koppel, M., Pennebaker, J. W., & Schler, J. (2009). Automatically profiling the author of an anonymous text. *Communications of the ACM, 52*(2), 119–123. https://doi.org/10.1145/1461928.1461959

Ashcroft, M., Kaati, L., & Meyer, M. (2015). A Step Towards Detecting Online Grooming – Identifying Adults Pretending to be Children. *Proceedings of the 2015 European Intelligence and Security Informatics Conference*, 98–104. https://doi.org/10.1109/EISIC.2015.41

Bark (2022). *Bark Parental Controls*. Bark. https://www.bark.us/ (retrieved December 7, 2022).

Barnett, J., & Coulson, M. (2010). Virtually Real: A Psychological Perspective on Massively Multiplayer Online Games. *Review of General Psychology, 14*(2), 167–179. https://doi.org/10.1037/a0019442

BBC (2013, July 11). 'Virtual Lolita' aims to trap chatroom paedophiles. *BBC News*. https://www.bbc.com/news/technology-23268893 (retrieved December 7, 2022).

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer. https://link.springer.com/book/9780387310732

Black, P. J., Wollis, M., Woodworth, M., & Hancock, J. T. (2015). A linguistic analysis of grooming strategies of online child sex offenders: Implications for our understanding of predatory sexual behavior in an increasingly computer-mediated world. *Child Abuse & Neglect, 44*, 140–149. https://doi.org/10.1016/j.chiabu.2014.12.004

Blei, D. M. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Bogdanova, D., Rosso, P., & Solorio, T. (2014). Exploring high-level features for detecting cyberpedophilia. *Computer Speech and Language, 28*(1), 108–120. https://doi.org/10.1016/j.csl.2013.04.007

Borj, P. R., & Bours, P. (2019). Predatory Conversation Detection. *Proceedings of the 2019 International Conference on Cyber Security for Emerging Technologies (CSET)*, 1–6. https://doi.org/10.1109/CSET.2019.8904885

Borj, P. R., Raja, K., & Bours, P. (2020). On Preprocessing the Data for Improving Sexual Predator Detection. *Proceedings of the 2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA)*, 1–6. https://doi.org/10.1109/SMAP49528.2020.9248461

Bours, P., & Kulsrud, H. (2019). Detection of Cyber Grooming in Online Conversation. *Proceedings of the 2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, 1–6. https://doi.org/10.1109/WIFS47025.2019.9035090

Briggs, P., Simon, W. T., & Simonsen, S. (2011). An Exploratory Study of Internet-Initiated Sexual Offenses and the Chat Room Sex Offender: Has the Internet Enabled a New Typology of Sex Offender? *Sexual Abuse: A Journal of Research and Treatment, 23*(1), 72–91. https://doi.org/10.1177/1079063210384275

Bundeskriminalamt (2022, February 18). *T05 Grundtabelle – Straftaten mit Tatmittel „Internet" – Fallentwicklung (V1.0)*. PKS 2021 Bund Falltabellen. https://www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/PolizeilicheKriminalstatistik/PKS2021/PKSTabellen/BundFalltabellen/bundfalltabellen.html?nn=194190 (retrieved December 16, 2022).

Callejas-Rodríguez, Á., Villatoro-Tello, E., Meza, I., & Ramírez-de-la-Rosa, G. (2016). From Dialogue Corpora to Dialogue Systems: Generating a Chatbot with Teenager Personality for Preventing Cyber-Pedophilia. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Proceedings of the 19th International Conference on Text, Speech, and Dialogue* (Vol. 9924, pp. 531–539). Springer International Publishing. https://doi.org/10.1007/978-3-319-45510-5 61

Cano, A. E., Fernandez, M., & Alani, H. (2014). Detecting Child Grooming Behaviour Patterns on Social Media. In L. M. Aiello & D. McFarland (Eds.), *Proceedings of the 6th International Conference on Social Informatics (SocInfo 2014)* (Vol. 8851, pp. 412–427). Springer International Publishing. https://doi.org/10.1007/978-3-319-13734-6_30

Cardei, C., & Rebedea, T. (2017). Detecting sexual predators in chats using behavioral features and imbalanced learning. *Natural Language Engineering, 23*(4), 589–616. https://doi.org/10.1017/S1351324916000395

Carvalho, J., & Nobre, P. J. (2019). Five-Factor Model of Personality and Sexual Aggression. *International Journal of Offender Therapy and Comparative Criminology, 63*(5), 797–814. https://doi.org/10.1177/0306624X13481941

Chapman, C., & Stolee, K. T. (2016). Exploring regular expression usage and context in Python. *ISSTA 2016: Proceedings of the 25th International Symposium on Software Testing and Analysis*, 282–293. https://doi.org/10.1145/2931037.2931073

Cheong, Y.-G., Jensen, A. K., Guðnadóttir, E. R., Bae, B.-C., & Togelius, J. (2015a). Detecting Predatory Behavior in Game Chats. *IEEE Transactions on Computational Intelligence and AI in Games, 7*(3), 220–232. https://doi.org/10.1109/TCIAIG.2015.2424932

Dalal, M. K., & Zaveri, M. A. (2011). Automatic Text Classification: A Technical Review. *International Journal of Computer Applications, 28*(2), 37–40. https://doi.org/10.5120/3358-4633

Dhouioui, Z., & Akaichi, J. (2016). Privacy Protection Protocol in Social Networks Based on Sexual Predators Detection. *ICC '16: Proceedings of the International Conference on Internet of Things and Cloud Computing, 63*, 1–6. https://doi.org/10.1145/2896387.2896448

Ebrahimi, M., Suen, C. Y., & Ormandjieva, O. (2016). Detecting predatory conversations in social media by deep Convolutional Neural Networks. *Digital Investigation, 18*(C), 33–49. https://doi.org/10.1016/j.diin.2016.07.001

Escalante, H. J., Montes y Gomez, M., Villasenor, L., & Errecalde, M. L. (2016). Early text classification: A Naïve solution. *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 91–99. https://doi.org/10.18653/v1/W16-0416

Escalante, H. J., Villatoro-Tello, E., Garza, S. E., López-Monroy, A. P., Montes-y-Gómez, M., & Villaseñor-Pineda, L. (2017). Early detection of deception and aggressiveness using profile-based representations. *Expert Systems with Applications, 89*, 99–111. https://doi.org/10.1016/j.eswa.2017.07.040

Faraz, A., Mounsef, J., Raza, A., & Willis, S. (2022). Child Safety and Protection in the Online Gaming Ecosystem. *IEEE Access, 10*, 115895–115913. https://doi.org/10.1109/ACCESS.2022.3218415

Fauzi, M. A., & Bours, P. (2020). Ensemble Method for Sexual Predators Identification in Online Chats. *Proceedings of the 2020 8th International Workshop on Biometrics and Forensics (IWBF)*, 1–6. https://doi.org/10.1109/IWBF49977.2020.9107945

Grant, T., & MacLeod, N. (2016). Assuming Identities Online: Experimental Linguistics Applied to the Policing of Online Paedophile Activity. *Applied Linguistics, 37*(1), 50–70. https://doi.org/10.1093/applin/amv079

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies, 43*(5–6), 907–928. https://doi.org/10.1006/ijhc.1995.1081

Gunawan, F. E., Ashianti, L., Candra, S., & Soewito, B. (2016). Detecting online child grooming conversation. *Proceedings of the 2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, 1–6. https://doi.org/10.1109/KICSS.2016.7951413

Gupta, A., Kumaraguru, P., & Sureka, A. (2012). *Characterizing Pedophile Conversations on the Internet using Online Grooming* (arXiv:1208.4324). arXiv; arXiv e-prints. https://doi.org/10.48550/arXiv.1208.4324

Harms, C., & Ferlazzo, M. (2007, July 12). *ISU prof provides online predator identification tips in Sex Offender Law Report*. Iowa State University – News Service. https://www.news.iastate.edu/news/2007/jul/predators.shtml (retrieved December 14, 2022).

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 28*(1), 100–108. https://doi.org/10.2307/2346830

Hidalgo, J. M. G., & Díaz, A. A. C. (2012). Combining Predation Heuristics and Chat-Like Features in Sexual Predator Identification. *CLEF (Online Working Notes/Labs/Workshop)*, 1–6. https://pan.webis.de/downloads/publications/papers/gomezhidalgo_2012.pdf

HMIC (2014). *An inspection of undercover policing in England and Wales*. https://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/an-inspection-of-undercover-policing-in-england-and-wales.pdf (retrieved November 30, 2022).

Inches, G., & Crestani, F. (2012). Overview of the International Sexual Predator Identification Competition at PAN-2012. *CLEF (Online Working Notes/Labs/Workshop)*, 1–12. https://pan.webis.de/downloads/publications/papers/inches_2012.pdf

InfoSec (2022). *G.A.R.S | UoM InfoSec Research Group*. https://infosec.uom.gr/?page_id=1391 (retrieved December 7, 2022).

Jacovi, A., Sar Shalom, O., & Goldberg, Y. (2018). Understanding Convolutional Neural Networks for Text Classification. *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 56–65. https://doi.org/10.18653/v1/W18-5408

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation, 28*(1), 11–21. https://doi.org/10.1108/eb026526

Kent, S. (2013, July 22). Virtual Lolita 'Negobot' targets pedophiles. *Ottawa Sun*. https://ottawasun.com/2013/07/22/virtual-lolita-negobot-targets-pedophiles (retrieved December 7, 2022).

Kim, J., Kim, Y. J., Behzadi, M., & Harris, I. G. (2020). Analysis of Online Conversations to Detect Cyberpredators Using Recurrent Neural Networks. *Proceedings of the Workshop on Social Threats in Online Conversations: Understanding and Management (STOC-2020)*, 15–20. https://par.nsf.gov/servlets/purl/10167942

Kodžoman, V., Marče, P., & Škaro, A. (2016). Sexual Predator Identification Using Ensemble Learning Classifiers. *Text Analysis and Retrieval 2016: Course Project Reports (TAR 2016)*, 32–35.

Kontostathis, A., Edwards, L., & Leatherman, A. (2010). Text Mining and CyberCrime. In M. W. Berry & J. Kogan (Eds.), *Text Mining: Applications and Theory* (pp. 149–164). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470689646.ch8

Kontostathis, A., West, W., Garron, A., Reynolds, K., & Edwards, L. (2012). Identifying Predators Using ChatCoder 2.0. *CLEF (Online Working Notes/Labs/Workshop)*, 1–10.

Lantz, B. (2019). *Machine Learning with R: Expert techniques for predictive modeling* (03 ed.). Packt Publishing Ltd.

Laorden, C., Galán-García, P., Santos, I., Sanz, B., Hidalgo, J. M. G., & Bringas, P. G. (2012). Negobot: A Conversational Agent Based on Game Theory for the Detection of Paedophile Behaviour. *Proceedings of the International Joint Conference CISIS'12-ICEUTE 12-SOCO 12 Special Sessions*, 261–270.

López-Monroy, A. P., González, F. A., Montes-y-Gómez, M., Escalante, H. J., & Solorio, T. (2018). Early Text Classification Using Multi-Resolution Concept Representations. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1216–1225. https://doi.org/10.18653/v1/N18-1110

MacFarlane, K., & Holmes, V. (2016). Multi-agent System for Safeguarding Children Online. *Proceedings of SAI Intelligent Systems Conference (IntelliSys), 2*, 228–242. https://doi.org/10.1007/978-3-319-56991-8_18

MacLeod, N., & Grant, T. (2016). "You have ruined this entire experiment…shall we stop talking now?" Orientations to the experimental setting as an interactional resource. *Discourse, Context & Media, 14*, 63–70. https://doi.org/10.1016/j.dcm.2016.10.001

MacLeod, N., & Grant, T. (2017). "go on cam but dnt be dirty": Linguistic levels of identity assumption in undercover online operations against child sex abusers. *Language and Law / Linguagem e Direito, 4*(2), 157–175.

MacLeod, N., & Wright, D. (2020). Forensic linguistics. In S. Adolphs & D. Knight (Eds.), *The Routledge Handbook of English Language and Digital Humanities* (1st ed., pp. 360–377). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781003031758-19/forensic-linguistics-nicci-macleod-david-wright (retrieved August 11, 2023).

McGhee, I., Bayzick, J., Kontostathis, A., Edwards, L., McBride, A., & Jakubowski, E. (2011). Learning to Identify Internet Sexual Predation. *International Journal of Electronic Commerce, 15*(3), 103–122. https://doi.org/10.2753/JEC1086-4415150305

Meyers, R. A. (Ed.) (2000). Clustering and Classification of Analytical Data. In *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation* (pp. 9689–9709). John Wiley & Sons, Ltd.

Michalopoulos, D., Mavridis, I., & Jankovic, M. (2014). GARS: Real-time system for identification, assessment and control of cyber grooming attacks. *Computers & Security, 42*, 177–190. https://doi.org/10.1016/j.cose.2013.12.004

Milon-Flores, D. F., & Cordeiro, R. L. F. (2022). How to take advantage of behavioral features for the early detection of grooming in online conversations. *Knowledge-Based Systems, 240*(C), 1–41. https://doi.org/10.1016/j.knosys.2021.108017

Misra, K., Devarapalli, H., Ringenberg, T. R., & Rayz, J. T. (2019). Authorship Analysis of Online Predatory Conversations using Character Level Convolution Neural Networks. *Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 623–628. https://doi.org/10.1109/SMC.2019.8914323

Mladenović, M., Ošmjanski, V., & Stanković, S. V. (2021). Cyber-aggression, Cyberbullying, and Cyber-grooming: A Survey and Research Challenges. *ACM Computing Surveys (CSUR), 54*(1), 1–42. https://doi.org/10.1145/3424246

Morris, C., & Hirst, G. (2012). Identifying Sexual Predators by SVM Classification with Lexical and Behavioral Features. *CLEF (Online Working Notes/Labs/Workshop)*, 1–12. https://pan.webis.de/downloads/publications/papers/morris_2012.pdf

Muñoz, F., Isaza, G., & Castillo, L. (2020). SMARTSEC4COP: Smart Cyber-Grooming Detection Using Natural Language Processing and Convolutional Neural Networks. *Proceedings of the 17th International Symposium on Distributed Computing and Artificial Intelligence*, 11–20. https://doi.org/10.1007/978-3-030-53036-5_2

Ngejane, C. H., Eloff, J. H. P., Sefara, T. J., & Marivate, V. N. (2021). Digital forensics supported by machine learning for the detection of online sexual predatory chats. *Forensic Science International: Digital Investigation, 36*, 1–37. https://doi.org/10.1016/j.fsidi.2021.301109

Ngejane, C. H., Mabuza-Hocquet, G., Eloff, J. H. P., & Lefophane, S. (2018). Mitigating Online Sexual Grooming Cybercrime on Social Media Using Machine Learning: A Desktop Survey. *Proceedings of the 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (IcABCD)*, 1–6. https://doi.org/10.1109/ICABCD.2018.8465413

Nijman, H., Merckelbach, H., & Cima, M. (2009). Performance intelligence, sexual offending and psychopathy. *Journal of Sexual Aggression, 15*(3), 319–330. https://doi.org/10.1080/13552600903195057

NTNU (2022). AiBA – Real-time, continuous, multimodal detection of sexual predators online. *NTNU Technology Transfer AS*. https://www.ntnutto.no/prosjekter-items/aiba/ (retrieved December 8, 2022).

O'Connell, R. (2003). *A typology of child cybersexploitation and online grooming practices*. Cyberspace Research Unit, University of Central Lancashire. http://image.guardian.co.uk/sys-files/Society/documents/2003/07/17/Groomingreport.pdf

Pandey, S. J., Klapaftis, I., & Manandhar, S. (2012). Detecting Predatory Behaviour from Online Textual Chats. In A. Dziech & A. Czyżewski (Eds.), *Proceedings of the International Conference on Multimedia Communications, Services and Security* (pp. 270–281). Springer Science & Business Media. https://doi.org/10.1007/978-3-642-30721-8_27

Parapar, J., Losada, D. E., & Barreiro, Á. (2014). Combining Psycho-linguistic, Content-based and Chat-based Features to Detect Predation in Chatrooms. *Journal of Universal Computer Science, 20*(2), 213–239.

Peersman, C., Daelemans, W., & Van Vaerenbergh, L. (2011). Predicting Age and Gender in Online Social Networks. *SMUC'11: Proceedings of the 3rd International Workshop on Search and Mining User-Generated Contents*, 37–44. https://doi.org/10.1145/2065023.2065035

Peersman, C., Vaassen, F., Van Asch, V., & Daelemans, W. (2012). Conversation Level Constraints on Pedophile Detection in Chat Rooms. *CLEF (Online Working Notes/Labs/Workshop)*, 1–13. https://pan.webis.de/downloads/publications/papers/peersman_2012.pdf

Pendar, N. (2007). Toward Spotting the Pedophile Telling victim from predator in text chats. *Proceedings of the International Conference on Semantic Computing (ICSC 2007)*, 235–241. https://doi.org/10.1109/ICSC.2007.32

Penna, L., Clark, A., & Mohay, G. (2010). A Framework for Improved Adolescent and Child Safety in MMOs. *Proceedings of the 2010 International Conference on Advances in Social Networks Analysis and Mining*, 33–40. https://doi.org/10.1109/ASONAM.2010.66

Penna, L., Clark, A., & Mohay, G. (2013). Enhancing Child Safety in MMOGs. In T. Özyer, J. Rokne, G. Wagner, & A. H. P. Reuser (Eds.), *The Influence of Technology on Social Network Analysis and Mining* (pp. 471–495). Springer Science+Business Media. https://doi.org/10.1007/978-3-7091-1346-2_21

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The Development and Psychometric Properties of LIWC2015*. The University of Texas at Austin. https://repositories.lib.utexas.edu/handle/2152/31333

Perverted Justice Foundation (2022). *Perverted Justice*. http://www.perverted-justice.com/ (retrieved December 7, 2022).

Popescu, M., & Grozea, C. (2012). Kernel Methods and String Kernels for Authorship Analysis. *CLEF (Online Working Notes/Labs/Workshop)*, 1–12. https://pan.webis.de/downloads/publications/papers/popescu_2012.pdf

Preuß, S., Bley, L. P., Bayha, T., Dehne, V., Jordan, A., Reimann, S., Roberto, F., Zahm, J. R., Siewerts, H., Labudde, D., & Spranger, M. (2021). Automatically Identifying Online Grooming Chats Using CNN-based Feature Extraction. *Proceedings of the 17th Conference on Natural Language Processing (KONVENS), 17*, 137–146. https://aclanthology.org/2021.konvens-1.12

Qustodio LLC (2022). *Die All-in-One-Lösung für Kindersicherung und digitales Wohlbefinden*. Qustodio. https://www.qustodio.com/de/ (retrieved December 8, 2022).

Raffel, L., Bours, P., & Komandur, S. (2020). Attention! Designing a Target Group-Oriented Risk Communication Strategy. In C. Stephanidis & M. Antona (Eds.), *Proceedings of the 22nd International Conference on Human-Computer Interaction* (pp. 597–604). Springer Nature. https://doi.org/10.1007/978-3-030-50732-9_77

Razi, A., Kim, S., Soubai, A., Stringhini, G., Solorio, T., De Choudhury, M., & Wisniewski, P. J. (2021). A Human-Centered Systematic Literature Review of the Computational Approaches for Online Sexual Risk Detection. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW2), 1–38. https://doi.org/10.1145/3479609

Ringenberg, T., Misra, K., Seigfried-Spellar, K. C., & Rayz, J. T. (2019). Exploring Automatic Identification of Fantasy-Driven and Contact-Driven Sexual Solicitors. *Proceedings of the 2019 Third IEEE International Conference on Robotic Computing (IRC)*, 532–537. https://doi.org/10.1109/IRC.2019.00110

Salton, G. (1968). *Automatic Information Organization and Retrieval.* McGraw Hill Text.

Sapkota, U., Bethard, S., Montes-y-Gómez, M., & Solorio, T. (2015). Not All Character N-grams Are Created Equal: A Study in Authorship Attribution. *Proceedings of the 2015 Conference of the North American Chapter of the ACL: Human Language Technologies*, 93–102. https://doi.org/10.3115/v1/N15-1010

Seigfried-Spellar, K. C. (2018). Assessing the Psychological Well-being and Coping Mechanisms of Law Enforcement Investigators vs. Digital Forensic Examiners of Child Pornography Investigations. *Journal of Police and Criminal Psychology, 33*(3), 215–226. https://doi.org/10.1007/s11896-017-9248-7

Shao, Y., Taylor, S., Marshall, N., Morioka, C., & Zeng-Treitler, Q. (2018). Clinical Text Classification with Word Embedding Features vs. Bag-of-Words Features. *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data)*, 2874–2878. https://doi.org/10.1109/BigData.2018.8622345

Siva, K., Baskar, A., Ramesh, A., Rengarajan, G., Shanmugam, G., Selvabharathi, S., & Sangeetha, D. (2021). Prevention of Emotional Entrapment of Children on Social Media. *Proceedings of the 2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)*, 95–100. https://doi.org/10.1109/ICETCI51973.2021.9574068

Sunde, N., & Sunde, I. M. (2021). Conceptualizing an AI-based Police Robot for Preventing Online Child Sexual Exploitation and Abuse: *Nordic Journal of Studies in Policing, 8*(2), 1–21. https://doi.org/10.18261/issn.2703-7045-2021-02-01

Tomljanović, J., Zuanović, L., & Šebrek, T. (2016). Sexual Predator Identification Using word2vec Features. *Text Analysis and Retrieval 2016: Course Project Reports (TAR 2016)*, 70–72.

Toriumi, F., Nakanishi, T., Tashiro, M., & Eguchi, K. (2015). Analysis of User Behavior on Private Chat System. *Proceedings of the 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 3*, 1–4. https://doi.org/10.1109/WI-IAT.2015.49

Tyagi, A. K., & Rekha, G. (2019, February). Machine Learning with Big Data. *Proceedings of International Conference on Sustainable Computing in Science, Technology & Management (SUSCOM)*. International Conference on Sustainable Computing in Science, Technology & Management (SUSCOM-2019), Jaipur, India. https://doi.org/10.2139/ssrn.3356269

van de Loo, J., De Pauw, G., & Daelemans, W. (2016). Text-Based Age and Gender Prediction for Online Safety Monitoring. *International Journal of Cyber-Security and Digital Forensics, 5*(1), 46–60. https://doi.org/10.17781/P002012

Vartapetiance, A., & Gillam, L. (2012). Quite Simple Approaches for Authorship Attribution, Intrinsic Plagiarism Detection and Sexual Predator Identification. *Notebook for PAN at CLEF 2012*. CLEF (Online Working Notes/Labs/Workshop), Rome, Italy.

Vilariño, D., Castillo, E., Pinto, D., Olmos, I., & León, S. (2012). Information Retrieval and Classification based Approaches for the Sexual Predator Identification. *CLEF (Online Working Notes/Labs/Workshop)*, 1–4. https://pan.webis.de/downloads/publications/papers/vilarino_2012.pdf

Villatoro-Tello, E., Juárez-González, A., Escalante, H. J., Montes-y-Gómez, M., & Villaseñor-Pineda, L. (2012). A Two-step Approach for Effective Detection of Misbehaving Users in Chats. *CLEF (Online Working Notes/Labs/Workshop)*, 1–12. https://pan.webis.de/downloads/publications/papers/villatorotello_2012.pdf

Wani, M. A., Agarwal, N., & Bours, P. (2021). Sexual-predator Detection System based on Social Behavior Biometric (SSB) Features. *Procedia Computer Science, 189*, 116–127. https://doi.org/10.1016/j.procs.2021.05.075

Webis Group (2022). *PAN*. https://pan.webis.de/ (retrieved December 2, 2022).

Winters, G. M., & Jeglic, E. L. (2022). Online Sexual Grooming. In *Sexual Grooming. Integrating Research, Practice, Prevention, and Policy* (pp. 65–86). Springer Nature. https://doi.org/10.1007/978-3-031-07222-2

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation, 1*(1), 67–82. https://doi.org/10.1109/4235.585893

Wondershare. (2022). *The Most Reliable Parental Control App*. FamiSafe. https://famisafe.wondershare.com/ (retrieved December 7, 2022).

Yahyavi, A., & Kemme, B. (2013). Peer-to-peer architectures for massively multiplayer online games: A Survey. *ACM Computing Surveys, 46*(1), 1–51. https://doi.org/10.1145/2522968.2522977

Yasaka, K., Akai, H., Kunimatsu, A., Kiryu, S., & Abe, O. (2018). Deep learning with convolutional neural network in radiology. *Japanese Journal of Radiology, 36*(4), 257–272. https://doi.org/10.1007/s11604-018-0726-3

Zambrano, P., Torres, J., Tello-Oquendo, L., Jacome, R., Benalcazar, M. E., Andrade, R., & Fuertes, W. (2019). Technical Mapping of the Grooming Anatomy Using Machine Learning Paradigms: An Information Security Approach. *IEEE Access, 7*, 142129–142146. https://doi.org/10.1109/ACCESS.2019.2942805

Zhai, C., & Massung, S. (2016). *Text Data Management and Analysis. A Practical Introduction to Information Retrieval and Text Mining*. ACM and Morgan & Claypool Publishers.

Ziegler, E. (2004). "Die Grenzen meiner Tastatur sind die Grenzen meiner Pseudonymkonstruktion": Form und Funktion von Chat-Pseudonymen im IRC. *Bulletin VALS-ASLA, 80*, 109–123.

Zuo, Z., Li, J., Anderson, P., Yang, L., & Naik, N. (2018). Grooming Detection using Fuzzy-Rough Feature Selection and Text Classification. *Proceedings of the 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–8. https://doi.org/10.1109/FUZZ-IEEE.2018.8491591