

Algorithms can prevent online abuse

August 24 2022, by Anne-Lise Aakervik and Nancy Bazilchuk



Cyber grooming: An adult male can pretend to be a 14-year-old boy online. Now NTNU detection technology in use at a commercial spin-off called AiBA, can detect the perpetrator. Credit: AiBA home page

Millions of children log into chat rooms every day to talk with other children. One of these "children" could well be a man pretending to be a 12-year-old girl with far more sinister intentions than having a chat about "My Little Pony" episodes.

Inventor and NTNU professor Patrick Bours at AiBA is working to

prevent just this type of predatory behavior. AiBA, an AI-digital moderator that Bours helped found, can offer a tool based on behavioral biometrics and algorithms that detect sexual abusers in online chats with children.

And now, as recently reported by Dagens Næringsliv, a national financial newspaper, the company has raised capital of NOK 7.5. million, with investors including Firda and Wiski Capital, two Norwegian-based firms.

In its newest efforts, the company is working with 50 million chat lines to develop a tool that will find high-risk conversations where abusers try to come into contact with children. The goal is to identify distinctive features in what abusers leave behind on gaming platforms and in social media.

"We are targeting the major game producers and hope to get a few hundred games on the platform," Hege Tokerud, co-founder and general manager, told Dagens Næringsliv.

Cyber grooming a growing problem

Cyber grooming is when adults befriend children online, often using a fake profile.

However, "some sexual predators just come right out and ask if the child is interested in chatting with an older person, so there's no need for a fake identity," Bours said.

The perpetrator's purpose is often to lure the children onto a private channel so that the children can send pictures of themselves, with and without clothes, and perhaps eventually arrange to meet the young person.

The perpetrators don't care as much about sending pictures of themselves, Bours said. "Exhibitionism is only a small part of their motivation," he said. "Getting pictures is far more interesting for them, and not just still pictures, but live pictures via a webcam."

"Overseeing all these conversations to prevent abuse from happening is impossible for moderators who monitor the system manually. What's needed is automation that notifies moderators of ongoing conversation," says Bours.

AiBA has developed a system using several algorithms that offers large chat companies a tool that can discern whether adults or children are chatting. This is where behavioral biometrics come in.

An [adult male](#) can pretend to be a 14-year-old boy online. But the way he writes—such as his typing rhythm, or his choice of words—can reveal that he is an adult man.

Machine learning key

The AiBA tool uses machine learning methods to analyze all the chats and assess the risk based on certain criteria. The risk level might go up and down a little during the conversation as the system assesses each message. The red warning symbol lights up the chat if the risk level gets too high, notifying the moderator who can then look at the conversation and assess it further.

In this way, the algorithms can detect conversations that should be checked while they are underway, rather than afterwards when the damage or abuse might have already occurred. The algorithms thus serve as a warning signal.

Cold and cynical

Bours analyzed loads of chat conversations from old logs to develop the [algorithm](#).

"By analyzing these conversations, we learn how such men 'groom' the recipients with compliments, gifts and other flattery, so that they reveal more and more. It's cold, cynical and carefully planned," he says.

"Reviewing chats is also a part of the learning process such that we can improve the AI and make it react better in the future."

"The danger of this kind of contact ending in an assault is high, especially if the abuser sends the recipient over to other platforms with video, for example. In a live situation, the algorithm would mark this chat as one that needs to be monitored."

Analysis in real time

"The aim is to expose an abuser as quickly as possible," says Bours.

"If we wait for the entire [conversation](#) to end, and the chatters have already made agreements, it could be too late. The monitor can also tell the child in the chat that they're talking to an adult and not another child."

AiBA has been collaborating with gaming companies to install the algorithm and is working with a Danish game and chat platform called MoviestarPlanet, which is aimed at [children](#) and has 100 million players.

In developing the algorithms, the researcher found that users write differently on different platforms such as Snapchat and TikTok.

"We have to take these distinctions into account when we train the algorithm. The same with language. The service has to be developed for all types of language," says Bours.

Looking at chat patterns

Most recently, Bours and his colleagues have been looking at chat patterns to see which patterns deviate from what would be considered normal.

"We have analyzed the chat patterns—instead of texts—from 2.5 million chats, and have been able to find multiple cases of grooming that would not have been detected otherwise," Bours said.

"This initial research looked at the data in retrospect, but currently we are investigating how we can use this in a system that follows such chat patterns directly and can make immediate decisions to report a user to a moderator," he said.

Provided by Norwegian University of Science and Technology

Citation: Algorithms can prevent online abuse (2022, August 24) retrieved 27 April 2024 from <https://techxplore.com/news/2022-08-algorithms-online-abuse.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.