European Commission

Artificial Intelligence in the fight against online CSAM

(Highlights Report)

INHOPE

How can Artificial Intelligence be used by Internet Hotlines to identify and remove Child Sexual Abuse Material on the Internet?

On Safer Internet Day 2020, 11 February 2020, an EU-funded INHOPE Focus Group took place exploring the topic of how Artificial Intelligence (AI) can be used by Internet Hotlines and law enforcement agencies (LEAs) to identify and remove Child Sexual Abuse Material (CSAM) on the internet instantly and at source. A key objective of the Focus Group was to explore the role of Internet Hotlines in the use of AI to fight online CSAM and to take stock of existing technologies, their capabilities and faults. The Focus Group also aimed to understand the needs of Internet Hotlines and LEAs, and to identify where automation could help in their work. Moreover, the Focus Group aimed to explore the development of a universal standard which would allow investigators and Internet Hotlines worldwide to exchange and access all data. CSAM is a crime where all of the evidence is in front of our eyes – the challenge currently faced is to draw together the different pieces which are spread throughout a gigantic mass of data and which is still collected in too many silo databases.

What Is Artificial Intelligence?

AI is a technology solution that can assure a more effective and efficient detection and takedown of online Child Sexual Abuse Material (CSAM). AI has started to play a prominent role in the discovery and ultimate removal of CSAM by helping to recognise, categorise and triage material. Law enforcement agencies (LEAs) and technology companies are already developing enhanced technologies including AI to recognise CSAM.

Artificial Intelligence is the philosophical and practical idea that machines can carry out tasks in a way that humans consider 'smart,' by humans programming them to fulfil these needs. Originally developing AI was a question of making computers perform challenging sequential tasks, like calculating increasingly complex calculations. However, as technology and our understanding of the human brain changed, work on AI has been directed towards mimicking human decision-making processes and carrying out tasks in more 'human' ways. It now relies on the adage that intelligence is learning without instruction or programming. The current cutting edge of AI is machine learning (ML), which is born from the idea that machines should be able to train on data and learn for themselves.

Use of Artificial Intelligence in relation to fighting online CSAM

AI is already being used in the fight against CSAM by service and social media providers who use analytics and AI to detect and stop exploitation within their networks. An example is Google's Content Safety API, a tool that uses Artificial Intelligence to help organisations better prioritise CSAM for review. These tools are offered for free to NGOs and private companies to protect children. Google, Facebook and other tech companies have staff who investigate and gather information from reports provided from the public, as well as their own automated systems which are then followed up by humans. These methods allow cases to be created and passed on to the police when necessary.

Law enforcement agencies have started to use AI to reduce the time spent by officers assessing and prioritising CSAM reports. Tools use AI techniques to investigate the content of imagery so that they can prioritise reports for human review. The AviaTor project, with EU-funding from the Internal Security Fund, aims at developing automation and intelligence tools to reduce the time spent by law enforcement officers in processing, assessing and prioritising reports on CSAM enabling them to spend more time on investigations and victim identification.

Under the Connecting Europe Facility (CEF), the European Commission is currently funding the secure online platform ICCAM that allows for a swift exchange of reports between 47 Internet Hotlines in 43 countries worldwide. ICCAM does not use AI. However, especially as it advances, AI technology is one of the best tools we have at our fingertips to fight CSAM.

Besides the different types of technology being used today, including hashing technology, blocking and filter technologies, web crawlers, and keyword matching, AI should be an additional step to tackling CSAM. It can be used to help process reports faster and avoid duplication at the same time as assuring more accuracy with the objective of faster take down of seen and previously unseen CSAM.

Key Recommendations

The meeting on Safer Internet Day 2020 brought together INHOPE Member Hotline representatives from European countries, the US and Canada, as well as stakeholders from industry, security and technology experts, international law enforcement agencies and academics. The aim of the day was to reflect on an action plan on the use of AI and the role it can play in the future needs of hotlines and law enforcement.

Main takeaways

Artificial Intelligence can help with the processing of CSAM. The main challenge lies with the quality of CSAM data and the lack of standardised classifications among the different parties involved. Involved NGOs, Hotlines, law enforcement agencies and the internet industry all currently classify content in a different way. However, to train machines one needs large quantities of data. If the data from different stakeholders varies (in its classification), then each party will have to train their own systems which reduces efficiency in eliminating CSAM from the internet. Thus, the effective use of AI and ML in the fight against online CSAM firstly requires a focus on standardised classification of material.

The following recommendations were put forward for INHOPE, the European Commission, national law enforcement agencies and the internet industry while all parties continue to develop technology that uses AI to fight CSAM.

Specific Recommendations to the use of Artificial Intelligence

- Common standards, taxonomies and classifications of online CSAM should be created;

- INHOPE should take the lead in identifying a unified and standardised classification of CSAM which is considered illegal across i) all EU, ii) all member Hotlines with the purpose of having data which can be used for machine learning;

- A review and needs assessment should be conducted among international law enforcement agencies and INHOPE Internet Hotlines to identify what is achievable and practical in terms of machine learning with current systems;

- INHOPE, in close concert with INTERPOL and national law enforcement agencies, should develop a concept of data types to be used in machine learning among Hotlines and should implement and roll this out for ICCAM;

- Many of the barriers in using AI are 'soft' rather than technical and therefore a campaign to raise awareness of this to relevant stakeholders would create impact and gain more support and collaboration, for finances and staffing;

- Public authorities that are allowed to have the data in order to train the models are needed (like AviaTor is doing with the Dutch police today) to which every country in the EU (or the INHOPE members) should send their data. The resulting models could then be shared with Hotlines (possibility to set up a geographical source on which the data can be trained).

Overall recommendations:

• There needs to be a bigger emphasis on prevention, especially with behavioural trends and impact;

• There should be a (technology) platform where more mature companies share their practices with start-ups / younger companies incentivising companies through grant apps, funding for safety tech start-ups, and trying to get the controllers of the funding (VC - Venture Capital) to value safety. Safety by design is an issue tech companies need to take seriously;

• An assessment of the impact of privacy data matters on using large data sets for machine learning should be conducted as a collaborative effort between the industry, Hotlines and law enforcement agencies.

• Training prosecutors and judges is essential to ensure they understand recorded crimes.

• There should be agreement with ISPs and companies, that any images with a context relating to abuse images should also be removed.

For more Information on this report, please email communications@Inhope.org

Created: March 2020