Statement Martina Tschapka, Director Operations and Content Manager Online Child Safety bei t3k.ai

PRESSEKONFERENZ zur Präsentation der Ergebnisse der Umfrage „Help us to help you", Befragung von Usern von Kindesmissbrauchsdarstellungen (Child Sexual Abuse Material – CSAM) direkt im weltweiten Darknet / 20.4.2023 via Zoom

Ich möchte zu Beginn kurz auf die unterschiedlichen Technologien zur Erkennung von CSAM eingehen. Mittels Hashes von Dateien werden eindeutige Matches gefunden, die zuvor validiert und in Datenbanken gespeichert wurden, hier gibt es keine False Positives, außer es gibt falsche Einträge in Datenbanken, was durch die Validierung im EU Zentrum vermieden wird. PhotoDNA kann auch leicht veränderte Dateien erkennen, ist daher robuster in der Erkennung, aber auch etwas anfälliger für False Positives. Mit üblichen Settings sind diese aber auch gut zu handhaben. Auch PhotoDNA Matching benötigt Datenbanken validierter bekannter Inhalte als Grundlage. KI lernt auch neue Inhalte zu erkennen, die bislang nicht in Datenbanken enthalten sind. KI wird kontinuierlich weiter trainiert, wir selbst arbeiten dazu in einem neuartigen Ansatz auch mit abstrahierten Daten, um das Training zu verbessern. Wenn sie trainiert ist, erkennt KI immer bestimmte Muster, seien es visuelle Inhalte, die auf Kindesmissbrauch hindeuten, oder bspw. Gesprächsmuster im Fall von Grooming. Oftmals wird leider von Erkennungsraten und False Positives gesprochen, ohne darauf Bezug zu nehmen, dass die Präzision von KI mittels gesetzter Schwellwerte stark beeinflusst werden kann.

Diese unterschiedlichen Technologien sollen laut dem Entwurf der EU Richtlinie, nach erfolgten und geprüften Detection Orders, sofern sie sich als qualitativ und proportional angemessen, sowie auch als sicher und maximal datenschutzwahrend im Einsatz erwiesen haben, vom neuen EU Zentrum zur Verfügung gestellt und von betroffenen Unternehmen angewandt werden. Dem EU Zentrum und dessen dediziertem unabhängigem Technologiekomitee fällt dabei auch schon in einem Prozess zuvor die Rolle zu, die Technologien auf Funktionalität und Sicherheit zu prüfen, und entsprechende Sicherheitsstandards zu implementieren.

Die Technologien können je nach Anwendungsfall und Art der Datenweitergabe implementiert werden, um vorliegende und hochgeladene Daten eines Onlinespeichers oder sozialen Mediums zu scannen, oder aber zum Beispiel vor Versand einer Nachricht. In Datenbanken vorliegende verifizierte illegale Dateien, sowie Treffer von KI Lösungen, werden angezeigt und für manuelles Review aufbereitet, um Fehler auszuschließen. Nutzer werden vor Umsetzung über diesen Schritt informiert.

Der erste Schritt zur Detection Order ist eine individuelle Risikoeinschätzung der Unternehmen, auf Basis deren unterschiedlich abgestufte Maßnahmen zu treffen sein werden. In einigen Fällen wird es dabei wohl auch im Falle einer Detection Order bei der Erkennung von bekanntem Material mittels Hashabgleich bleiben. Allerdings ist KI immer dann notwendig, wenn es um Material geht, das noch nicht bekannt ist. Gerade hinter diesem Begriff verbirgt sich oftmals aktueller Missbrauch, den es zu erkennen und zu unterbrechen gilt. Und nur durch Erkennung von neuem Material kann es auch aktualisierte Datenbanken von bekanntem Material geben.

Alle oben beschriebenen Technologien sind bereits erfolgreich im Einsatz, zumeist auf offenen Plattformen. Ein großes Thema ist auch die automatisierte Erkennung von Indikatoren in verschlüsselter Kommunikation, um wenn sie vorliegen, anlassbezogen kontrollieren und einschreiten zu können. Im Fall des Grooming gibt es gute Ansätze, es wird aber noch Arbeit notwendig sein. Dass diese notwendig ist, zeigt auch die präsentierte Studie.

Die EU Richtlinie strebt aber jedenfalls nicht danach, Verschlüsselung aufzuheben, diese ist ein wichtiger Pfeiler und wird das auch bleiben. Angestrebt wird die Prüfung existierender Tools und Technologien, sowie deren Weiterentwicklung, um Privacy und Sicherheit nicht gegeneinander abzuwiegen, sondern nebeneinanderzustellen.

Abschließend und zusammenfassend möchte ich drei Punkte betonen:

- Die Kombination von Technologien wird immer notwendig sein und umgesetzt werden – um Fehler zu minimieren und Erkennung zu maximieren. KI wird nicht als alleiniges Mittel eingesetzt werden, sondern in Ergänzung der Erkennung von bekanntem Material.

- Adaptierungen und Spezifizierungen im Entwurf der Richtlinie werden notwendig sein, und die Technologien müssen gründlich evaluiert und gegebenenfalls weiterentwickelt werden, aber es ist wichtig, eine Richtlinie umzusetzen.

- Der vorliegende Entwurf fußt auf dem Grundsatz technologischer Neutralität, um mit den Entwicklungen auf Seiten der Unternehmen, aber auch Trends auf Seiten der Kriminellen Schritt halten zu können, und so zum Kindeswohl beizutragen

Kontakt:
Martina Tschapka
Director Operations | Content Manager Online Child Safety t3k.ai
Martina.tschapka@t3k.at