Offener Brief zur Machbarkeit von “Chat Control”: Einschätzungen aus wissenschaftlicher Sicht

Der unter dem Namen “Chat Control” diskutierte Vorstoß der EU-Kommission, anlasslose Überwachung verschiedener Kommunikationskanäle zur Erkennung von kinderpornographischem, terroristischem oder sonstigem “unerwünschten” Material – bis hin zum Versuch der Früherkennung (z.B. “Grooming” Minderjähriger durch Vertrauen aufbauende Textnachrichten) – verpflichtend für Mobilgeräte und Kommunikationsdienste einzuführen, wurde kürzlich um die Überwachung direkter Audiokommunikation erweitert. Einige Staaten, unter anderem Österreich und Deutschland, haben bereits öffentlich erklärt, diesen Vorstoß der anlassunabhängigen Überwachung nicht zu unterstützen. Auch Zivilschutz- sowie Kinderrechtsorganisationen haben dieses Vorgehen als überzogen und gleichzeitig ineffektiv abgelehnt. Jüngst wurde selbst vom juristischen Dienst des EU-Ministerrats eine Unvereinbarkeit mit den europäischen Grundrechten diagnostiziert. Ungeachtet dessen wird der Entwurf sogar noch verschärft und auf weitere Kanäle ausgeweitet: in der letzten Ausprägung sogar auf Audionachrichten und Gespräche. Das Vorgehen scheint mit entsprechenden Versuchen in den USA (“EARN IT” und “STOP CSAM” Acts) und dem UK (“Online Safety Bill”) koordiniert zu sein.

Als Wissenschaftler:innen, die aktiv in verschiedenen Bereichen dieser Thematik forschen, geben wir daher in aller Klarheit die Erklärung ab: Dieser Vorstoß ist nicht sicher und effektiv umsetzbar. Derzeit ist keine Weiterentwicklung der entsprechenden Technologien absehbar, die eine solche Umsetzung technisch ermöglichen würden. Zudem sind unserer Einschätzung nach die erhofften Effekte dieser Überwachungsmaßnahmen nicht zu erwarten. Diese Gesetzesinitiative verfehlt daher ihr Ziel, ist gesellschaftspolitisch gefährlich und würde die Sicherheit unserer Kommunikationskanäle für den größten Teil der Bevölkerung nachhaltig schädigen.

Die Hauptgründe gegen die Machbarkeit von “Chat Control” wurden bereits mehrfach genannt. Wir möchten diese im Folgenden speziell in der interdisziplinären Verbindung zwischen **Artificial Intelligence** (AI, Künstliche Intelligenz / KI), **Security** (Informationssicherheit / Technischer Datenschutz) und **Recht** erörtern.

Unsere Bedenken sind:

1. Security: a) Verschlüsselung ist die beste Methode für Internet-Sicherheit. Gelungene Angriffe liegen fast immer an fehlerhafter Software. b) Eine systematische und automatisierte Überwachung (d.h. “Scannen”) verschlüsselter Inhalte ist technisch nur möglich, wenn die durch Verschlüsselung erzielbare Sicherheit massiv verletzt wird, was mit erheblichen zusätzlichen Risiken einhergeht. c) Eine gesetzliche Verpflichtung zur Integration solcher Scanner wird sichere digitale Kommunikation in der EU für den Großteil der Bevölkerung unverfügbar machen, aber kaum Auswirkung auf kriminelle Kommunikation haben.
2. AI: a) Automatisierte Klassifizierung von Inhalten, u.a. mit auf Machine Learning basierten Methoden, ist immer mit Fehlern behaftet, die in diesem Fall zu hohen Falschmeldungen führen werden. b) Speziell Überwachungsmethoden, die auf den Endgeräten ausgeführt werden, eröffnen noch zusätzliche Angriffsmöglichkeiten bis hin zur Extraktion des u.U. illegalen Trainingsmaterials.
3. Recht: a) Eine vernünftige Abgrenzung zu explizit erlaubter Verwendung spezifischer Inhalte z.B. im Bildungsbereich oder für Kritik und Parodie scheint automatisiert nicht möglich. b) Der massive Grundrechtseingriff durch ein solches Instrument der Massenüberwachung ist nicht verhältnismäßig und würde große Kollateralschäden in der Gesellschaft erzeugen.

Im Detail gründen sich diese Bedenken auf folgende wissenschaftlich anerkannte Tatsachen:

1. Security
    a. Verschlüsselung mit modernen Methoden ist eine unverzichtbare Grundlage praktisch aller technischen Mechanismen zur Wahrung von Sicherheit und Datenschutz im Internet. Auf diese Weise wird derzeit die Kommunikation im Internet als Grundstein für aktuelle Dienste bis hin zu kritischer Infrastruktur wie Telefon-, Strom-, Wassernetze, Krankenhäuser usw. geschützt. Das Vertrauen in gute Verschlüsselungsmethoden ist unter Experten deutlich höher als in andere Sicherheitsmechanismen. Vor allem die durchschnittlich mangelhafte Qualität von Software im Allgemeinen ist der Grund für die vielen auch öffentlich bekannten Sicherheitsvorfälle. Eine Verbesserung dieser Situation im Sinne besserer Sicherheit stützt sich daher primär auf Verschlüsselung. Eine generelle Verbesserung der Software-Qualität und damit vernünftige Sicherheit für nicht-verschlüsselte Inhalte ist nicht absehbar.
    b. Eine automatische Überwachung (“Scannen”) korrekt verschlüsselter Inhalte ist nach aktuellem Stand des Wissens nicht effektiv möglich. Verfahren wie “Fully Homomorphic Encryption” (FHE) sind für diesen Einsatz derzeit nicht geeignet - weder ist das Verfahren dazu in der Lage, noch ist die notwendige Rechenleistung realistisch verfügbar. Eine schnelle Verbesserung ist auch hier nicht absehbar.
    c. Aus diesen Gründen wurde von früheren Versuchen, Ende-zu-Ende-Verschlüsselung zu verbieten oder einzuschränken, international zumeist rasch wieder abgegangen. Der aktuelle Chat-Control-Vorstoß zielt darauf ab, Überwachungsfunktionalität in der Form von Scanning-Modulen in die Endgeräte (*“Client-Side Scanning” / CSS*) einbauen zu lassen und daher vor der sicheren Verschlüsselung oder nach der sicheren Entschlüsselung die Klartextinhalte zu scannen.
    Anbieter von Kommunikationsdiensten müssten gesetzlich verpflichtet werden, dies für alle Inhalte umzusetzen. Da dies nicht im Kerninteresse solcher Organisationen liegt und Aufwand in Implementierung und Betrieb sowie technisch erhöhte Komplexität erfordert, ist von Freiwilligkeit bei der Einführung solcher Scanner nicht auszugehen - im Gegensatz zu Scanning auf Serverseite.
    d. Sichere Messenger wie z.B. Signal oder Threema und WhatsApp haben bereits öffentlich angekündigt, solche Client-Scanner nicht zu implementieren, sondern sich aus den entsprechenden Regionen zurückzuziehen. Dies hat verschiedene Auswirkungen auf Kommunikation je nach Anwendungsfall: (i) (Erwachsene) Kriminelle werden untereinander einfach über “nicht-konforme” Messenger-Dienste kommunizieren, um weiter von sicherer Verschlüsselung zu profitieren. Der erhöhte Aufwand, z.B. unter Android per Sideloading andere Apps, die im jeweiligen Land nicht über die üblichen App Stores erhältlich sind, zu installieren, ist für kriminelle Elemente keine nennenswerte Hürde. (ii) Kriminelle kommunizieren mit möglichen zukünftigen Opfern über populäre Plattformen, welche im Ziel der diskutierten verpflichtenden Überwachungsmaßnahmen wären. In diesem Fall ist davon auszugehen, dass informierte Kriminelle ihre Opfer schnell auf alternative aber dennoch international anerkannte Kanäle wie Signal locken, welche nicht von der Überwachung erfasst sind. (iii) Teilnehmer:innen tauschen problematisches Material aus, ohne sich der Tatsache bewusst zu sein, dass sie Straftaten begehen. Dieser Fall würde automatisch gemeldet und u.U. zur Kriminalisierung auch und vor allem Minderjähriger ohne Vorsatz führen. Von den Einschränkungen getroffen würde daher primär die breite - und unbescholtene - Masse der Bevölkerung.

Es wäre völlig illusorisch zu glauben, dass sichere Verschlüsselung ohne eingebaute Überwachung jetzt noch rückgängig gemacht werden könne. Tools wie Signal, Tor, Cwtch, Briar und viele andere sind als Open Source breit verfügbar und können einfach der zentralen Kontrolle entzogen werden. Die Kenntnis um sichere Verschlüsselung ist bereits Allgemeinwissen und kann nicht mehr zensiert werden. **Es gibt keine effektive Möglichkeit, den Einsatz sicherer Verschlüsselung ohne Client-Side-Scanning (CSS) auf technischer Ebene zu blockieren. Wenn Überwachungsmaßnahmen in Messengern vorgeschrieben werden, werden nur noch Kriminelle ihre Privatsphäre wahren, deren eigentliche Verbrechen schwerer wiegen als der Verstoß gegen den Überwachungszwang.**

e. Weiters schafft die durch vorgeschlagene Scanner-Module erzwungene komplexe Implementierung zusätzliche Sicherheitsprobleme, die derzeit nicht existieren. Einerseits stellt dies neue Software-Komponenten dar, die wiederum angreifbar sein werden. Andererseits nehmen die Chat-Control-Vorschläge durchgehend an, dass die Scanner-Module selbst vertraulich bleiben werden, da sie einerseits mit Inhalten trainiert würden, deren bloßer Besitz (eingebaut in die Messenger-App) bereits strafbar ist und andererseits einfach zum Testen von Umgehungsmethoden verwendet werden können. Es ist ebenfalls eine Illusion, dass solche Machine-Learning-Modelle oder andere Scanner-Module, die auf Milliarden von Endgeräten unter der Kontrolle von Endbenutzer:innen verteilt werden, je geheim gehalten werden können. Ein besonders relevantes Beispiel ist das von Apple im Sommer 2021 vorgestellte “NeuralHash”-Modul für CSAM-Erkennung, welches fast unmittelbar aus entsprechenden iOS-Versionen extrahiert wurde und damit offen verfügbar ist. Die Annahme von Chat-Control-Vorschlägen, dass diese Scanner-Module vertraulich gehalten werden könnten, ist daher vollkommen unbegründet und falsch. Hier sind entsprechende Datenlecks fast unvermeidbar.

2. Artificial Intelligence
    a. Wir müssen annehmen, dass Machine-Learning (ML)-Modelle auf Endgeräten prinzipiell nicht vollständig geheim gehalten werden können. Dies steht im Gegensatz zu Scanning auf Serverseite, wie aktuell rechtlich möglich und auch aktiv von verschiedenen Anbietern praktiziert, um solche Inhalte zu scannen, die nicht Ende-zu-Ende-verschlüsselt wurden. ML-Modelle auf Serverseite können mit aktuellem Stand der Technik vernünftig vor Auslesen geschützt werden und stehen weniger im Fokus dieser Betrachtung.
    b. Ein generelles Problem bei allen ML-basierten Filtern sind Falschklassifikationen, also dass bekannt “ungewünschtes” Material bei kleinen Änderungen nicht als solches erkannt wird (auch als “*false negative*” oder “*false non-match*” bezeichnet). Für Teile der Vorstöße ist derzeit nicht bekannt, wie ML-Modelle komplexes, unbekanntes Material mit wechselndem Kontext (z.B. “Grooming” in Text-Chats) auch nur mit annähernder Genauigkeit erkennen können sollten. Die Wahrscheinlichkeit hoher False-negative-Raten ist hoch.

        Im Sinne des Risikos deutlich schwerwiegender ist jedoch, wenn unbedenkliches Material als “unerwünscht” klassifiziert wird (auch als “*false positive*” oder “*false match*” oder auch als “collision” bezeichnet). Solche Fehler lassen sich reduzieren, aber prinzipiell nicht ausschließen. False-Positives führen neben der fälschlichen Beschuldigung unbeteiligter Personen zudem zu (u.U. sehr) vielen Falschmeldungen für die Ermittlungsbehörden, welche jetzt schon zu wenige Ressourcen haben, um Meldungen nachzugehen.
    c. Durch die anzunehmend offene Verfügbarkeit von ML-Modellen entstehen zudem verschiedene neue Angriffsmöglichkeiten. Am Beispiel Apple NeuralHash wurden sehr zeitnah zufällig auftretende Kollisionen gefunden und Programme frei veröffentlicht, um beliebige Kollisionen zwischen Bildern zu erzeugen. Dieses auch als “malicious collisions” bezeichnete Methode nützt sogenannte adversarial attacks

gegen das neuronale Netzwerk und ermöglicht so Angreifern, unbedenkliches Material bewusst als “match” im ML-Modell einstufen und damit als “unerwünscht” klassifizieren zu lassen. So können unschuldige Personen - ohne jegliche illegale Aktion auf Seiten der Angegriffenen oder Angreifer - gezielt durch automatische falsche Meldungen geschädigt und unter Verdacht gebracht werden.

d. Die offene Verfügbarkeit der Modelle kann auch für sogenanntes “training input recovery” verwendet werden, um aus dem ML-Modell die zum Training benutzten Inhalte (zumindest teilweise) zu extrahieren. Dies stellt bei verbotenen Inhalten (z.B. Kinderpornographie) ein weiteres massives Problem dar und kann den Schaden für Betroffene noch erhöhen, indem deren sensible Daten (z.B. Missbrauchsbilder, welche zum Training verwendet wurden) noch weiter veröffentlicht werden können. Aufgrund dieser und weiterer Probleme hat z.B. Apple den Vorschlag zurückgezogen. Wir halten fest, dass diese letztgenannte Gefahr bei serverseitigem Scanning durch ML-Modelle nicht auftritt, sondern neu durch den Chat-Control-Vorschlag mit Client-Scanner hinzukommt.

3. Rechtliche Aspekte
    a. Das Recht auf Privatsphäre ist ein Grundrecht, in das nur unter ganz engen Voraussetzungen eingegriffen werden darf. Wer dieses Grundrecht in Anspruch nimmt, darf nicht als von vornherein verdächtig gelten, etwas Kriminelles verbergen zu wollen. Der oft gebrauchte Satz: “Wer nichts zu verbergen hat, hat auch nichts zu befürchten!” verwehrt Menschen den Gebrauch ihres Grundrechts und fördert totalitäre Überwachungstendenzen. Der Einsatz von Chat Control würde dies befeuern.
    b. Insbesondere der Terrorismusbereich hat in seiner Breite Überschneidungen mit politischer Betätigung und freier Meinungsäußerung. Gerade vor diesem Hintergrund wird die “Vorfeldkriminalisierung”, wie sie in den vergangenen Jahren unter dem Deckmantel der Terrorismusbekämpfung vermehrt erfolgt ist, besonders kritisch gesehen. Chat-Control-Maßnahmen gehen in dieselbe Richtung. Sie können dieses Grundrecht stark beschneiden und Menschen, die sich politisch kritisch betätigen, in den Fokus krimineller Verfolgung rücken. Die dadurch ermöglichte starke Beschneidung politisch kritischer Betätigung hindert die Weiterentwicklung der Demokratie und birgt die Gefahr der Förderung radikalisierter Untergrundbewegungen.
    c. Zum Bereich von Rechts- und Sozialwissenschaften gehört es, kriminelle Phänomene zu erforschen und Regelungsmechanismen zu hinterfragen. Unter diesem Blickwinkel läuft auch wissenschaftlicher Diskurs Gefahr, im Wege von Chat Control als “verdächtig” identifiziert und damit mittelbar beschränkt zu werden. Die damit mögliche Stigmatisierung kritischer Rechts- und Sozialwissenschaften steht im Spannungsverhältnis zur Freiheit der Wissenschaften, welche auch “Forschung unabhängig von Mainstream” zur Weiterentwicklung benötigt.
    d. Im Bildungsbereich ist es erforderlich, junge Menschen zu kritischem Bewusstsein zu erziehen. Dazu gehört es etwa auch, Fakten über Terrorismus weiterzugeben. Durch den Einsatz von Chat Control könnte die Bereitstellung von Unterrichtsmaterial durch Lehrer:innen diese in einen kriminellen Fokus rücken. Gleiches gilt für die Thematisierung von sexuellem Missbrauch, sodass durch Kontrollmaßnahmen dieses sensible Thema, selbst wenn “Selbststärkungsmechanismen” gefördert werden sollen, im Ergebnis stärker tabuisiert werden könnte.
    e. Eingriffe in Grundrechte müssen, selbst wenn sie im Kontext der Strafverfolgung gesetzt werden, stets angemessen und verhältnismäßig sein. Die dargestellten technischen Überlegungen zeigen, dass diese Voraussetzungen bei Chat Control nicht gegeben sind. Damit fehlt es solchen Maßnahmen an jeglicher rechtlicher und ethischer Legitimität.

Zusammenfassend ist der aktuelle Vorschlag zur Chat-Control-Gesetzgebung weder aus Security- noch aus AI-Sicht technisch vernünftig und aus rechtlicher Sicht höchst problematisch und überbordend. Der Chat-Control-Vorstoß bringt deutlich größere Gefahren für die breite Bevölkerung als mögliche Verbesserung für Betroffene und ist daher abzulehnen.
Stattdessen sollten u.a. bestehende Möglichkeiten zur menschlich getriebenen Meldung möglicherweise problematischen Materials durch Empfänger:innen, wie durch verschiedene Messenger-Dienste bereits möglich, gestärkt und noch leichter zugänglich gemacht werden. Es ist zu überlegen, ob anonyme Einmeldemöglichkeiten für entsprechend illegales Material geschaffen und leicht von Messengern aus erreichbar gemacht werden könnten. Ebenfalls bereits bestehende Möglichkeiten der Strafverfolgung, wie z.B. durch Polizeibeamte durchgeführte Überwachung von sozialen Medien oder offenen Chatgruppen sowie die gesetzlich vorgesehene Analyse von Smartphones entsprechend Verdächtiger können weiterhin entsprechend genützt werden. Diese Verfahren lieferten auch bisher bereits verlässlichere Ergebnisse als eine durch Chat Control zu erwartende Überlast an automatischen Falschmeldungen, die durch Polizeiorgane verfolgt und einzeln als falsch identifiziert werden müssten.

Für ausführlichere Informationen und weitere Details stehen gerne zur Verfügung:

- Fragen der Security: Univ.-Prof. Dr. René Mayrhofer <rm@ins.jku.at>, Tel. 0732/2468-4121
- Fragen der AI: DI Dr. Bernhard Nessler <nessler@ml.jku.at>, Tel. 0732/2468-4489
- Fragen des Rechts: Univ.-Prof. Dr. Alois Birklbauer <alois.birklbauer@jku.at>, Tel 0732/2468-7447

Unterzeichner:innen

- **AI Austria, Verein zur Förderung von Künstlicher Intelligenz in Österreich**, Wollzeile 24/12, 1010 Wien
- **Austrian Society for Artificial Intelligence (ASAI),** Verein zur Förderung der wissenschaftlichen Forschung im Feld AI in Österreich
- Univ.-Prof. Dr. Alois Birklbauer, JKU Linz (**Leiter der Abteilung Praxis für Strafrechtswissenschaften und Medizinstrafrecht**)
- Ass.-Prof. Dr. Maria Eichlseder, TU Graz
- Univ.-Prof. Dr. Sepp Hochreiter, JKU Linz (**Vorstand des Instituts für Machine Learning, Leiter des LIT AI Labs**)
- Dr. Tobias Höller, JKU Linz (Post-doc am Institut für Netzwerke und Sicherheit)
- FH-Prof. DI. Peter Kieseberg, FH St. Pölten **(Leiter des Instituts für IT-Sicherheitsforschung)**
- Dr. Brigitte Krenn, Österreichisches Forschungsinstitut für Artificial Intelligence (**Board Member Austrian Society for Artificial Intelligence**)
- Univ.-Prof. Dr. Matteo Maffei, TU Wien (**Leiter des Security and Privacy Forschungsbereichs, Co-Leiter des TU Wien Cybersicherheitszentrums**)
- Univ.-Prof. Dr. Stefan Mangard, TU Graz (**Vorstand des Instituts für Angewandte Informationsverarbeitung und Kommunikationstechnologie)**
- Univ.-Prof. Dr. René Mayrhofer, JKU Linz (**Vorstand des Instituts für Netzwerke und Sicherheit**, **Co-Leiter des LIT Secure and Correct System Labs**)
- DI Dr. Bernhard Nessler, JKU Linz/SCCH (**Vizepräsident der Austrian Society for Artificial Intelligence**)
- Univ.-Prof. Dr. Christian Rechberger, TU Graz
- Dr. Michael Roland, JKU Linz (Post-doc am Institut für Netzwerke und Sicherheit)
- a.Univ.-Prof. Dr. Johannes Sametinger, JKU Linz (**Institut für Wirtschaftsinformatik - Software Engineering, LIT Secure and Correct System Labs**)
- Univ.-Prof. DI Georg Weissenbacher, DPhil (Oxon), TU Wien (Prof. f. Rigorous Systems Engineering)

Veröffentlicht am 4.7.2023