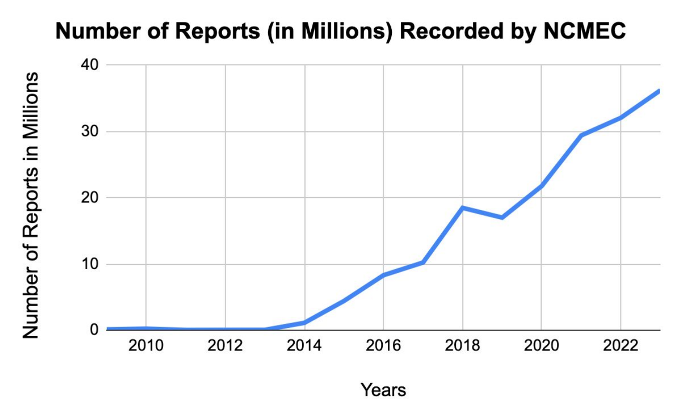
Child Sexual Abuse Material (CSAM)

About CSAM

- Images, videos, URLs, or text
- Usually: images
- Detection is driven by the National Center for Missing & Exploited Children (NCMEC)



About NCMEC

- Private, non-profit corporation
- "NCMEC is the nation's largest and most influential child protection organization."
- Centralized reporting of CSAM by companies and individuals

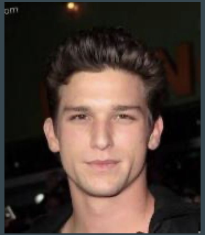
Position on End-to-End Encryption

- "Ignoring abuse won't stop it".
- "Without technological exceptions to end-to-end encryption, the dehumanizing abuse of children will continue"
- Urges tech companies to "enhance consumer privacy while <u>**prioritizing**</u> child safety."

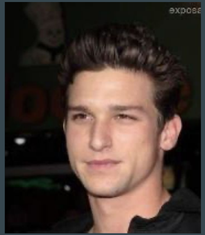
CSAM Detection

Perceptual hash functions principles

Similar images → Similar hashes

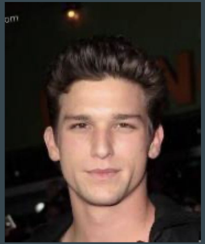


502ce0414f8e99c333f2f073



502ce2414fce99c333f2f073

Different images → Different hashes



502ce0414f8e99c333f2f073

05387734eeb66e5fa4b908eb

Detection process

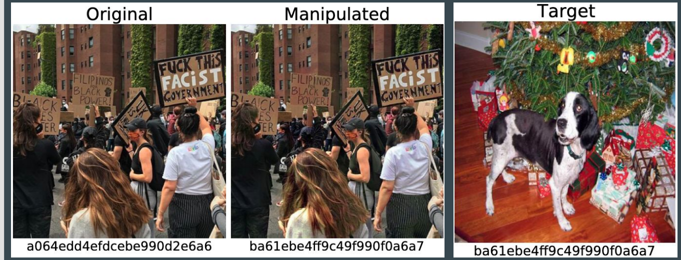
1. **Image Upload:**
   - The perceptual hash of the image is calculated.
   - The hash is compared against databases (e.g., NCMEC's 17.5M entries).
2. **Reporting:**
   - If a match is found, the image and user are reported to NCMEC.
3. **NCMEC Steps:**
   - **All reports are shared with law enforcement agencies through a tool co-developed by Meta and the U.S. government.**
   - Collaboration between NCMEC and these agencies.

Potential Privacy Concerns

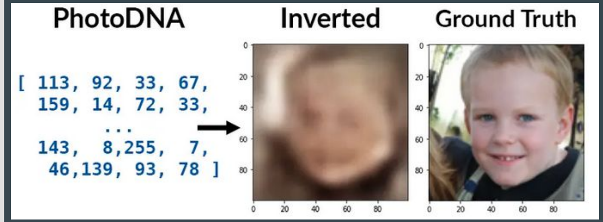
Challenges and Concerns

- **False Positives:** Innocent users could be falsely accused of possessing CSAM.
- **Client-Side Scanning:** Every photo and/or video on user devices is hashed.
- **Server Enclave:** The server has full access to all data.
- **Slippery Slope:** Allowing exceptions for CSAM could set a precedent for other exceptions.
- **Lack of Transparency:** The design, hashes, and parameters are not publicly disclosed.

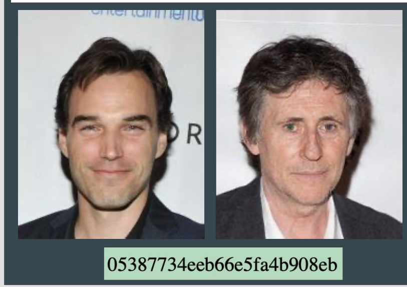
Attacks on Perceptual Hash Functions



Collision creation



Inversion



Natural Collisions

Current Status in Europe (ChatControl)

ChatControl Regulation

- 2021: Clear intention to define a regulation enabling the detection of CSAM
- 2021: Exception to the ePrivacy Regulation for CSAM detection
- 2022: First proposal rejected
- 2023: Second proposal rejected
- 2024: Third proposal not yet voted on, but discussions are ongoing
- 2024: Exception for CSAM detection extended to 2026

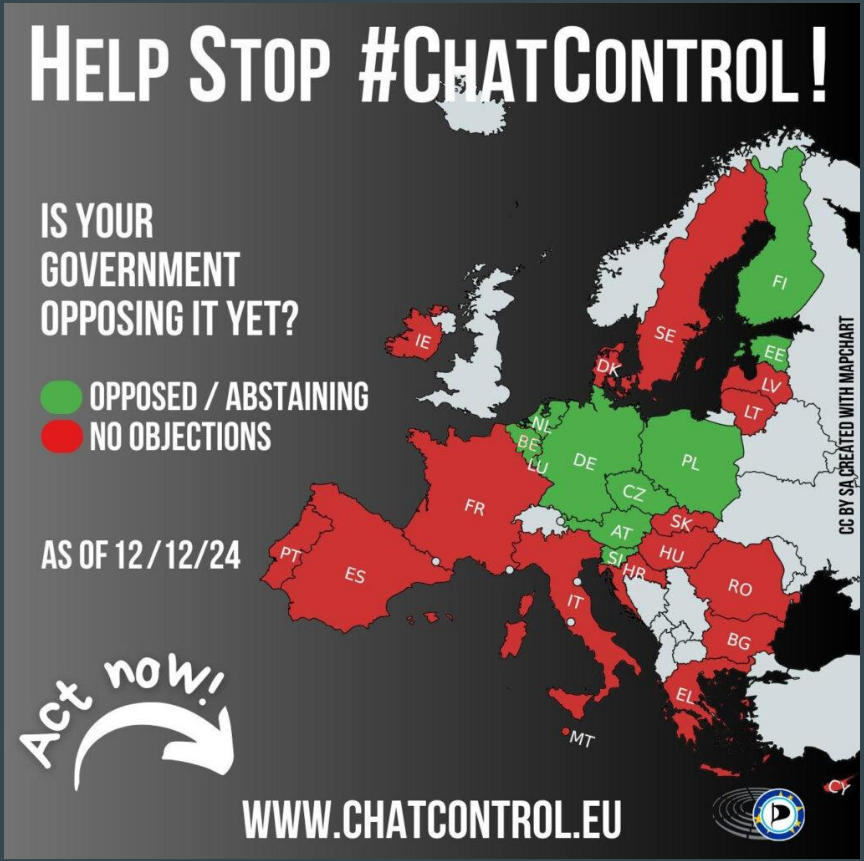
Key points

- Proposal for detecting CSAM in messages (images or videos) using External Service Providers (ESPs)
- Explicit mention of perceptual hash functions (e.g., PhotoDNA and NeuralHash)

Top 3 Proposed Solutions

- **Full Hashing:** Perceptual hash on device; matches checked on ESP server. If matched, app reports the user and content.
- **Partial Hashing:** Partial hash on device; same process as full hashing.
- **Server Enclave:** ESP server decrypts, checks for CSAM, and re-encrypts non-CSAM content.

12 December 2024 : EU Council discussed ChatControl proposal



Very Weak Blocking Minority:

In Favor: 17 governments, 60% of the population

Against: 10 governments, 40% of the population

Adoption Threshold: **15 governments** and **65%** of the population

Only 5% population missing for adoption

France currently supports ChatControl

All the details, summaries, and explanations about ChatControl:

www.chatcontrol.eu

Next Steps and Global Context

Europe and Beyond:

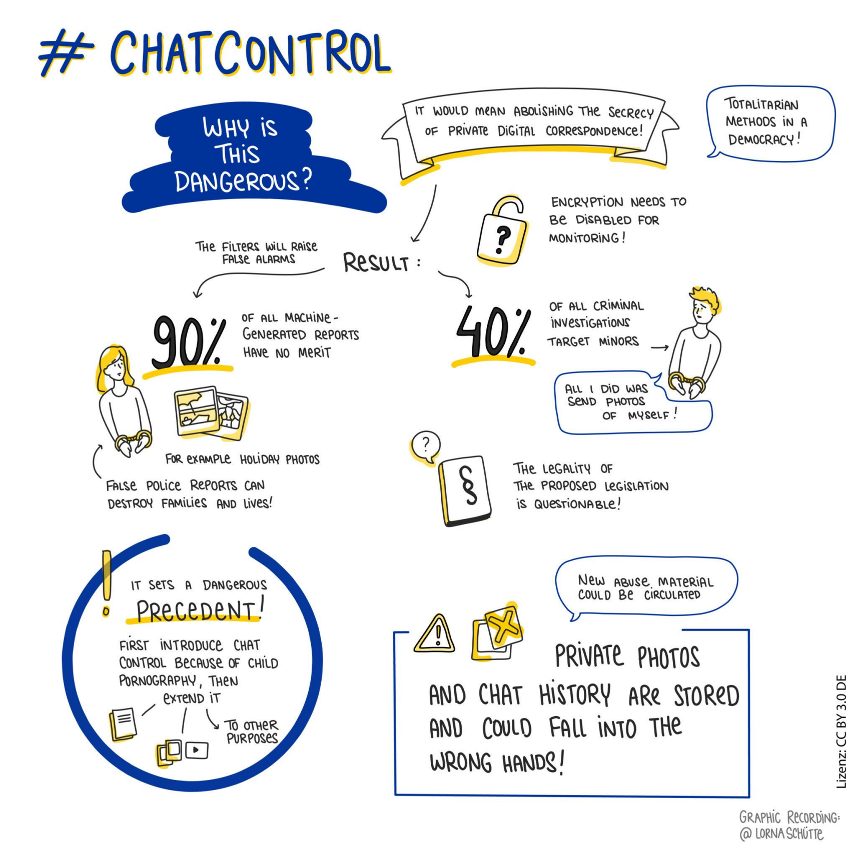
- New ChatControl regulation under discussion.
- Australia, Canada, New Zealand, UK, and USA signed the Voluntary Principles to Counter CSAM (2020) alongside tech companies.
- Growing intention to extend CSAM mechanisms beyond online uploaded content.
- UK: Passed the Online Safety Act (2023).
- Australia: Passed the "Cyber Security Act" (2024).

What we can do

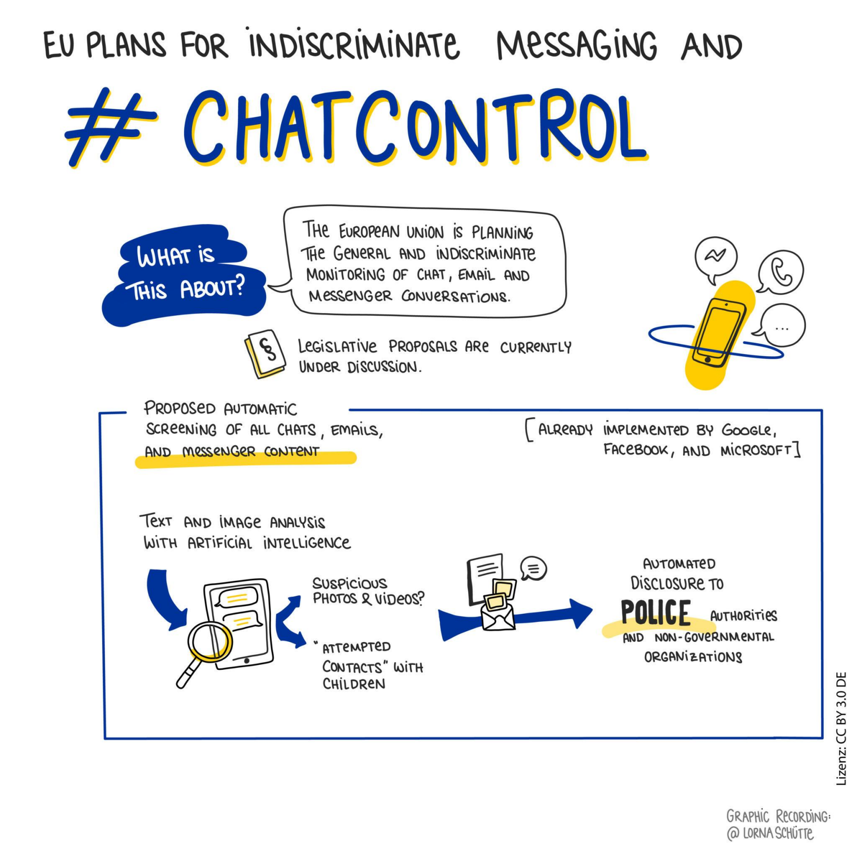
- **Sign open letters** advocating for privacy and safety
- **Discuss the issue** publicly
- Conduct research analyzing proposed solutions and their consequences
- Advocate for open design and transparency in perceptual hash functions used for CSAM detection
- **Defend end-to-end encryption**
- **Contact companies and policymakers** to warn them about potential risks
- Research alternative solutions to (CSS)
- Stop research on potentially harmful technologies

"Just because we can, doesn't mean we should."

Exemples communication


# CHATCONTROL
WHY IS THIS DANGEROUS?
IT WOULD MEAN ABOLISHING THE SECRECY OF PRIVATE DIGITAL CORRESPONDENCE!
TOTALITARIAN METHODS IN A DEMOCRACY!
RESULT:
ENCRYPTION NEEDS TO BE DISABLED FOR MONITORING!
THE FILTERS WILL RAISE FALSE ALARMS
90% OF ALL MACHINE-GENERATED REPORTS HAVE NO MERIT
40% OF ALL CRIMINAL INVESTIGATIONS TARGET MINORS
ALL I DID WAS SEND PHOTOS OF MYSELF!
FOR EXAMPLE HOLIDAY PHOTOS
FALSE POLICE REPORTS CAN DESTROY FAMILIES AND LIVES!
THE LEGALITY OF THE PROPOSED LEGISLATION IS QUESTIONABLE!
IT SETS A DANGEROUS PRECEDENT!
FIRST INTRODUCE CHAT CONTROL BECAUSE OF CHILD PORNOGRAPHY, THEN EXTEND IT
TO OTHER PURPOSES
NEW ABUSE MATERIAL COULD BE CIRCULATED
PRIVATE PHOTOS AND CHAT HISTORY ARE STORED AND COULD FALL INTO THE WRONG HANDS!
GRAPHIC RECORDING: @ LORNA SCHÜTTE

Lizenz: CC BY 3.0 DE


EU PLANS FOR INDISCRIMINATE MESSAGING AND
# CHATCONTROL
WHAT IS THIS ABOUT?
THE EUROPEAN UNION IS PLANNING THE GENERAL AND INDISCRIMINATE MONITORING OF CHAT, EMAIL AND MESSENGER CONVERSATIONS.
LEGISLATIVE PROPOSALS ARE CURRENTLY UNDER DISCUSSION.
PROPOSED AUTOMATIC SCREENING OF ALL CHATS, EMAILS, AND MESSENGER CONTENT
[ALREADY IMPLEMENTED BY GOOGLE, FACEBOOK, AND MICROSOFT]
TEXT AND IMAGE ANALYSIS WITH ARTIFICIAL INTELLIGENCE
SUSPICIOUS PHOTOS & VIDEOS?
"ATTEMPTED CONTACTS" WITH CHILDREN
AUTOMATED DISCLOSURE TO POLICE AUTHORITIES AND NON-GOVERNMENTAL ORGANIZATIONS
GRAPHIC RECORDING: @ LORNA SCHÜTTE

Lizenz: CC BY 3.0 DE

What Now?

- New regulations expected in the EU
- UK passed the “Online Safety Act”
- Australia passed the “ Cyber Security Act”

What about creation of perceptual hash Server Side Scanning function?

- Propose alternative solutions
- Oppose security through obscurity
- Research is not neutral: The end goal matters.
- "Just because we can, doesn't mean we should."
- Engage and inform policymakers and companies

What Now ?

Current Status in Europe (CHATCONTROL)

- Temporary exemption from the ePrivacy regulation since July 14, 2021
- ChatControl: Rejected in 2022 and 2023
- End of 2024: Exemption extended until 2026 + Search for a new consensus

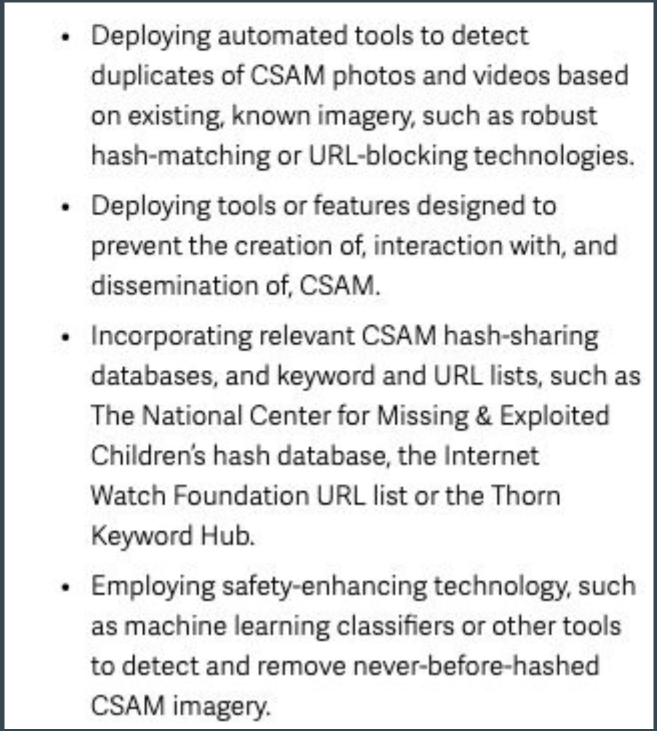
Key points:

- Preferred solutions: Client-Side Scanning (hash or partial hash)
- Explicit mention of perceptual hash functions

- « PhotoDNA has been in use for more than 10 years by over 150 organisations [...] and law enforcement in the EU[...]. In these 10 years, the tool has been used daily and analysed hundreds of billions of images without any accuracy concerns»
- "Rate of false positives is estimated at 1 in 50 billion" **Without sources** !

Elsewhere in the World

- Countries: US, Canada, UK, Australia, New Zealand
- **"Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse."**
- Signed by: Google, Apple, Amazon, Meta, Microsoft, Zoom, etc.
- Explicit mention of perceptual hash functions.



- Deploying automated tools to detect duplicates of CSAM photos and videos based on existing, known imagery, such as robust hash-matching or URL-blocking technologies.
- Deploying tools or features designed to prevent the creation of, interaction with, and dissemination of, CSAM.
- Incorporating relevant CSAM hash-sharing databases, and keyword and URL lists, such as The National Center for Missing & Exploited Children's hash database, the Internet Watch Foundation URL list or the Thorn Keyword Hub.
- Employing safety-enhancing technology, such as machine learning classifiers or other tools to detect and remove never-before-hashed CSAM imagery.