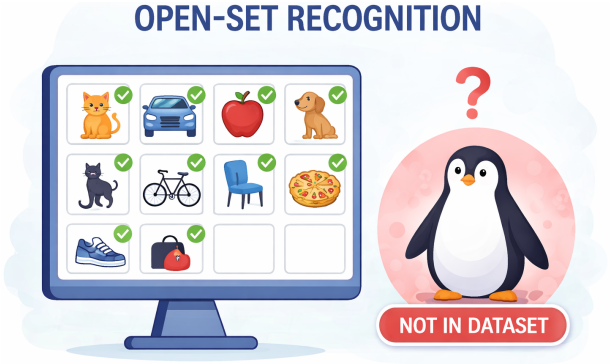
Secure Detection of Illegal Content



In 2022, the European Commission proposed a new regulation, often referred to as "chat control", aimed at combating online child sexual abuse material (CSAM) by scanning private messages and images shared on digital platforms (e.g., Facebook, Instagram, WhatsApp, Signal, etc.). As of now, platforms such as Facebook, Instagram, Google, and Microsoft already scan shared content to detect known CSAM, primarily using perceptual hash functions to compare images against CSAM hash databases.

Despite their widespread deployment, perceptual hash functions have been shown to be vulnerable to multiple attacks. These findings raise new research problems. Beyond the CSAM case, how can specific image classes be detected securely without compromising user privacy? Can sensitive content be detected reliably using existing techniques, and can approaches for computing on encrypted data (COED) address this problem? Moreover, current mechanisms focus on detecting whether sensitive content is present *in* a dataset, while detecting whether image classes are *not* present remains largely unexplored.

The intern will:

- Investigate techniques for detecting content that is unknown *a priori*, including perceptual hash functions, and AI-based methods such as open-set recognition.
- Assess adaptability of these methods for use with COED, and identify any required modifications or additional components.
- Study whether these techniques can be used to detect *the absence* of specific image classes (e.g. CSAM) in a given dataset, and evaluate their practicality in real-world scenarios.

High-level introduction to perceptual hash functions: `https://tsjournal.org/index.php/jots/article/view/24/14`
Examples of attacks on currently deployed functions: `https://www.usenix.org/system/files/usenixsecurity25-zhang-yushu.pdf`
AI-based open-set recognition techniques: `https://ieeexplore.ieee.org/abstract/document/9040673`
High-level introduction to COED techniques: `https://ieeexplore.ieee.org/document/10194492`

Practicalities

Promotor:	Bart Preneel		
Daily supervisors:	Mariana Gama	mariana.gama@esat.kuleuven.be	office B01.19
	Diane Leblanc-Albarel	diane.leblanc-albarel@esat.kuleuven.be	office B01.23
	Martin Zbudila	martin.zbudila@esat.kuleuven.be	office B01.19

 © 2025-2026 COSIC