Rethinking the detection of child abuse on the Internet

Elie Bursztein @elie
with the help of MANY collaborators
at NCMEC, Thorn and Google
WARNING
Potentially triggering language
and difficult subject ahead
Online child sexual abuse
The production, distribution, and consumption of digital child sexual abuse material (CSAM)
The online child sexual abuse ecosystem has transformed alongside technological shifts.
Child abusers take advantage of key technological shifts

- Camera phones easier to produce
- Rich media content diversification
- Online community easier to organize
- Digital distribution easier to share
We identified 5 key challenges through a multi-year data driven study.

To appear at WWW 2018
Let’s build technologies to create an Internet free of child abuse
NCMEC is the US clearinghouse for child abuse reports
Life of a CSAM report

The public

Report content

Online service

NCMEC

Refers

Feedback

Law enforcement
Some of the world’s clearinghouses

Canadian Center for child protection
~40K reports

NCMEC
18M reports

IWF
200K reports
The number of reports has grown exponentially.
1998
3,000 reports

2018
18,400,000 / 6000x reports
Content-based report prioritization to focus on the most actionable ones first
Open-source Tensorflow based system available through Thorn
An informative image

Gender:
- Female: 78.3%
- Male: 97.3%

Age Group:
- Young Adult: 63.3% (Female), 86.3% (Male)

Hair:
- Black: 87.1% (Female), 72.1% (Male)

Head Cover:
- None: 97.7% (Female), Cap: 62.1% (Male)
A less informative image

gender: male 100%
age group: young adult 92.3%
hair: unknown 74.4%
head cover: cap 97.7%
facial hair: beard 60.7%
eyewear: sunglasses 78.4%
A non informative image
Most content is only shared once
Key media types

84% Images reported once

91% Videos reported once
content very long tail makes traditional hash blacklists ineffective
Content aware detection to ensure new content is quickly identified and reported
Prioritization API

Decision logic

AI

Don’t review

Review

Toward applying AI to related images detection
Media types are rapidly diversifying
Video is the fastest growing medium
Video is the fastest growing medium
Reliable detection across all media to keep pace with medium diversification
YouTube video fingerprint
Child abuse cross borders
PlayPen dark web forum takedown

150,000
Forum users from around the world

548
International arrests

296
Abused children rescued

Streamline collaboration
to increase efficiency and
improve transparency
The emotional toll of manual review
Manual reviews are needed to check the validity of the reports and find potential clues that can help rescue childrens.
Reduce negative emotional impact via AI based image transformations
Style transfer explained

- Painting
- Image to review
- AI
- Image with painting style

Logos: Google, National Center for Missing & Exploited Children
Artistic style transfer *preserve image semantic while changing its appearance*
Highly experimental

Reduce negative emotion by 16% on explicit and violent images
Takeaway

- Technological shifts and weaponization
- Explosion of child sexual abuse content
- Build transformative technologies to curb online child abuse
Questions?
Let’s build together the technologies needed to make the Internet free of child abuse

contact: @elie / elieb@google.com