THE MOST IMPORTANT DATABASE YOU'VE NEVER HEARD OF

evelyn douek | edouek@law.harvard.edu | @evelyndouek

GIFCT

Global Internet Forum
to Counter Terrorism

2017

Founding Members

General Members

Government

Government of Canada, Public Safety Canada

Government of France, Ministry for Europe and Foreign Affairs

Government of Ghana, Ministry of Communications

Government of Japan, Ministry of Foreign Affairs

Government of New Zealand, Department of Prime Minister and Cabine

Government of the United Kingdom, Home Office

Government of the United States, National Security Council

CHRISTCHURCH CALL

TO ELIMINATE TERRORIST
& VIOLENT EXTREMIST
CONTENT ONLINE

Founders

- New Zealand
- France

Founding supporters announced in Paris, May 15 2019

- Australia
- Canada
- European Commission
- France
- Germany
- Indonesia
- India
- Ireland
- Italy
- Japan
- Jordan
- The Netherlands
- New Zealand
- Norway
- Senegal
- Spain
- Sweden
- United Kingdom

Supporters announced in New York on 23 September 2019

- Argentina
- Austria
- Belgium
- Bulgaria
- Chile
- Colombia
- Costa Rica
- Cyprus
- Denmark
- Finland
- Georgia
- Ghana
- Greece
- Hungary
- Iceland
- Ivory Coast
- Kenya
- Latvia
- Lithuania
- Luxembourg
- Maldives
- Malta
- Mexico
- Mongolia
- Poland
- Portugal
- Romania
- South Korea
- Slovenia
- Sri Lanka
- Switzerland
- UNESCO
- Council of Europe

To that end, we, the online service providers, commit to:

Work together to ensure cross-industry efforts are coordinated and robust, for instance by investing in and expanding the GIFCT, and by sharing knowledge and expertise.

To that end, we, Governments and online service providers, commit to work collectively to:

Support smaller platforms as they build capacity to remove terrorist and violent extremist content, including through sharing technical solutions and relevant databases of hashes or other relevant material, such as the GIFCT shared database.

Hash-Sharing Consortium

In 2016, the founding member companies of GIFCT (Facebook, Microsoft, Twitter, and YouTube), created a shared industry database of "hashes" — unique digital "fingerprints" — of known violent terrorist imagery or terrorist recruitment videos that had been removed from their services.

How does it work?

An image or video is "hashed" in its raw form and is not linked to any original platform or user data. Hashes appear as a numerical representation of the original content and cannot be reverse-engineered to recreate the image and/or video. A platform needs to find a match with a given hash on their platform in order to see what the hash corresponds with.

Members

The Hash-Sharing Consortium currently consists of 13 companies: *Microsoft, Facebook, Twitter, YouTube, Ask.fm, Cloudinary, Instagram, JustPaste.it, LinkedIn, Verizon Media, Reddit, Snap, and Yellow*

No access to non-industry members has been granted. Each consortium member can decide how they would like to use the database based on their own user terms of service, as well as how they operate and how they make use of both technical and human capabilities.

NUMBERS

- 2020 report: 250K visually distinct images and ~50K visually distinct videos (up from 200K overall in 2019)

BENEFITS

- Cross-platform threats ➔ ➔ ➔ ➔ Cross-platform responses
- Faster, more comprehensive response
- Technology sharing (esp. with smaller platforms)
- (?) Extra-legal take-downs (based on platforms' terms of service) are global

VOX Pol
ABOUT
BACKGROUND
EVENTS
Home / Blog / One Database to Rule Them All
The New York Times
YouTube Is Erasing History
Opinion
Under pressure to remove "extremist content," platforms are purging vital human rights evidence.
By Hadi Al Khatib and Dia Kayyali
Mr. Al Khatib is a human rights activist. Dia Kayyali is a program manager at WITNESS.
Oct. 23, 2019
expression.
ELECTRONIC FRONTIER FOUNDATION
A response to this article can be found HER
By Svea Windwehr and Jillian C York
The Invisible Content Cartel t
Online
Joint Letter to New Executive Director, Global Internet Forum to Counter Terrorism
HUMAN RIGHTS WATCH
CENTER FOR DEMOCRACY & TECHNOLOGY
embers of the European
ent on Concerns with Terrorism
Hash Database
February 4, 2019
Caught in the Net: The Impact of "Extremist" Speech Regulations on Human Rights Content
BY JILLIAN C. YORK | MAY 30, 2019

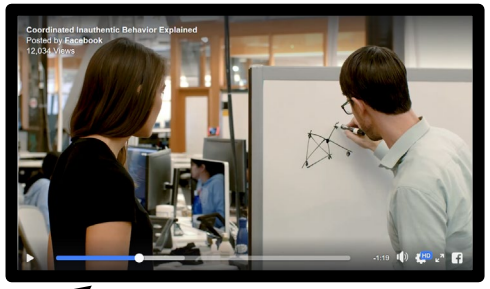
CONCERNS

- Lack of transparency—what is in the database??
- Power imbalances (both industry and gvt)
- Disparate impact of definition of “terrorist” and unclear definition of “glorification” (72% of the database)
- Overreliance on technical filters
- No remedial mechanisms
- Extra-legal take-downs (based on platforms’ terms of service) are <u>global!</u>

IMPORTANCE GOES BEYOND “TERRORIST” CONTENT

THE RISE OF CONTENT CARTELS ???

"CIB"/Influence operations



Terrorist Content

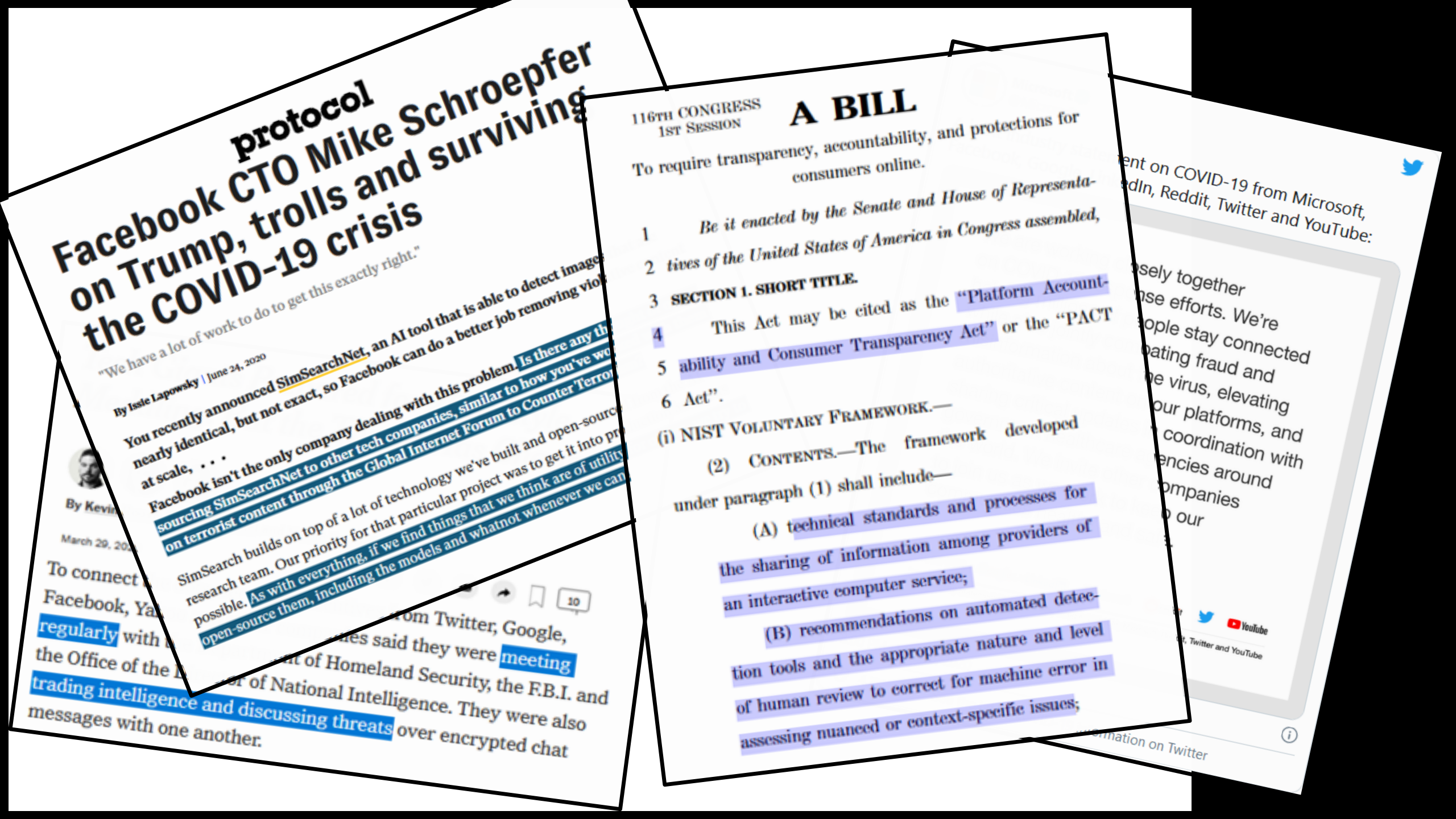
Global Internet Forum to Counter Terrorism

The Pilot

PhotoDNA™
The Next Chapter in Protecting Children Online

Frequency illusion or Baader–Meinhof phenomenon	The frequency illusion is that once something has been noticed then every instance of that thing is noticed, leading to the belief it has a high frequency of occurrence (a form of selection bias).[53] The Baader–Meinhof phenomenon is the illusion where something that has recently come to one's attention suddenly seems to appear with improbable frequency shortly afterwards.[54] The Baader–Meinhof phenomenon is sometimes conflated with frequency illusion and the recency illusion.[55] It was named after an incidence of frequency illusion in which the Baader–Meinhof Group was mentioned.[56]

https://en.wikipedia.org/wiki/List_of_cognitive_biases#Frequency_illusion



protocol

Facebook CTO Mike Schroepfer on Trump, trolls and surviving the COVID-19 crisis

"We have a lot of work to do to get this exactly right."

By Issie Lapowsky | June 24, 2020

You recently announced SimSearchNet, an AI tool that is able to detect image
nearly identical, but not exact, so Facebook can do a better job removing viol
at scale, . . .
Facebook isn't the only company dealing with this problem. Is there any th
sourcing SimSearchNet to other tech companies, similar to how you've wo
on terrorist content through the Global Internet Forum to Counter Terror

SimSearch builds on top of a lot of technology we've built and open-sourc
research team. Our priority for that particular project was to get it into pr
possible. As with everything, if we find things that we think are of utility
open-source them, including the models and whatnot whenever we can

10

By Kevin

March 29, 202

To connect
Facebook, Ya
regularly with
the Office of the D
trading intelligence and discussing threats over encrypted chat
messages with one another.

om Twitter, Google,
ies said they were meeting
t of Homeland Security, the F.B.I. and
r of National Intelligence. They were also

116TH CONGRESS
1ST SESSION

A BILL

To require transparency, accountability, and protections for consumers online.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the "Platform Accountability and Consumer Transparency Act" or the "PACT Act".

(i) NIST VOLUNTARY FRAMEWORK.—

(2) CONTENTS.—The framework developed under paragraph (1) shall include—

(A) technical standards and processes for the sharing of information among providers of an interactive computer service;

(B) recommendations on automated detection tools and the appropriate nature and level of human review to correct for machine error in assessing nuanced or context-specific issues;

ent on COVID-19 from Microsoft,
dIn, Reddit, Twitter and YouTube:

sely together
se efforts. We're
ople stay connected
ating fraud and
e virus, elevating
our platforms, and
coordination with
encies around
ompanies
our

YouTube
, Twitter and YouTube

mation on Twitter

CONCURRENT TRENDS

① Informal pressure

② Regulation that makes GIFCT effectively mandatory (e.g. EU TERREG)

③ Internet Referral Units (IRUs)

NEED TO BUILD IN:

- Transparency
- Oversight
- Remedial mechanisms for mistakes
- Guardrails on “creep”

THANK YOU

evelyn douek | edouek@law.harvard.edu | @evelyndouek