COUNTER
EXTREMISM
PROJECT

Thinking outside the box:
The truth about Social Media and implications
for the EU Digital Services Act (DSA)
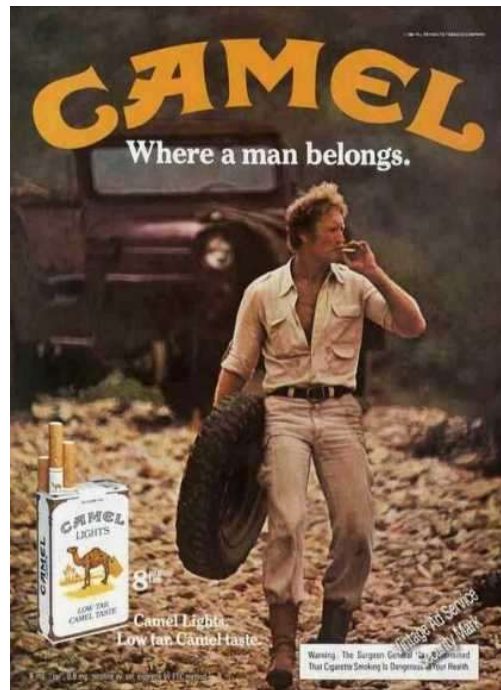
Alexander Ritzmann

14.07.2020

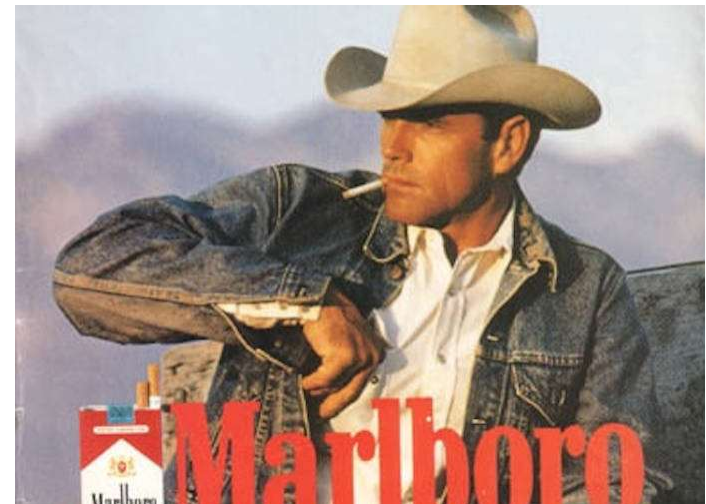# Let me … irritate you! *Why?*

There was a time…

Smoking is about…*Freedom*

Right?

- The movie follows the efforts of Big Tobacco's chief spokesman, Nick Naylor, who lobbies on behalf of cigarettes using heavy spin tactics…

A terrible misunderstanding!

- What are social media and video sharing platforms? I mean, **what are they really?**
- Social media **were never really intended to be "gardens of free speech".** Ask the investors.
- **Manipulation?** Projection? Fool me once, shame on you…fool me twice…
- They are just **businesses that sell access to theirs users data to third parties**. And they offer free services to those users. That is it. No magic. No gardens of free speech.

COUNTER EXTREMISM PROJECT

- Social media are **not public squares** but more like **virtual shopping malls** with big "speakers corners". **House-rules** trump freedom of expression. Try it at your local shopping center.
- Therefore: We need to treat and **regulate them like any other industry**, be it pharma, food or banking, **based on the potential harm** the services or products pose to EU citizens.
- The **"wunderkind" and "all-carrots-no-sticks"** era needs to come to an end.
- **Who told us** social media companies are about free speech and human rights?
- I wonder if there will be a movie about this, maybe called ***"Thank you for liking".***

NetzDG Stresstest

Lessons to be learned for the EU Digital _Services_ Act

CEP Berlin has carried out an **investigation** during February 2020. The aim was to **test** the extent to which YouTube, Facebook and Instagram block **"manifestly illegal" content** and labels of organizations **banned under German law**, following notification through the companies' respective NetzDG forms.

We also wanted to test the **"Theory of Change"** of "notice and take down" – **does it solve the problem at hand?**

COUNTER
EXTREMISM
PROJECT

Findings and lessons for the DSA

Out of the **92 pieces of manifestly illegal content** reported by CEP, 24 were blocked in accordance with NetzDG and 16 were deleted in accordance with platform guidelines. This corresponds to a **blocking / removal rate of 43.5%**.

On **YouTube**, the blocking / removal rate was **35%**. Videos with **identical content** were blocked in some cases and others not.

Facebook blocked the reported content but **did not block manifestly illegal content** found in the same folder.

Our reseach and assessmentes regarding **"illegal content"** were **100% correct.**

Does "notice and take down/action" work?
No, it doesent.

The aim of the "notice and take down" procedure prescribed *by* **the NetzDG and the DSA drafts** is to **make social media safer** for users. This can only succeed **if illegal content is found**, reported and blocked or removed effectively.

Currently, this procedure is largely **based on trust and chance** since content on the platforms is monitored on by the companies themselves, users and the Internet Referal Units (IRUs) of the police.

There is **no effective, systematic and continuous monitoring** of the platforms covered by the NetzDG in relation to violations of German laws. This means that **manifestly illegal content**, when "**unnoticed**" or **unreported**, can remain online **in large quantities.**

COUNTER
EXTREMISM
PROJECT

Automated systems - not IF but HOW?

Reservations against proactive measures and automated systems used by the platforms **are understandable**. It is a fact, however, that **companies are already using them against *unwanted* or *illegal*** content.

The question therefore is ***not IF* upload-filters** should be applied to prevent the dissemination of terrorist content online, ***but HOW*** to do that.

Regulation that focuses on transparency, auditability and **effectiveness would therefore protect civil liberties** more than no regulation.

COUNTER
EXTREMISM
PROJECT

We need <u>expainable</u> transparency and <u>capable</u> oversight

There is an urgent need for explainable transparency and auditability of processes and technologies that are used to implement content moderation policies.

The <u>SMCs themselves decide what they **can "see",**</u> what is or is not being removed and are **not required to be transparent** about the relevant figures and processes. Therefore it is possible for SMCs to claim that they *remove or block 99% of illegal content while **illegal content remains abundant** on those very same platforms.*

Our study, like many others, raises **doubts** that companies' **reports of success correspond with reality**.

EDiMA: Responsibility Online-restricted to broad measures

Which other industry is allowed to self-audit and self-report?

COUNTER
EXTREMISM
PROJECT

Proactive search for illegal content

The procedural logic of "notice and take down", on which the NetzDG is based, requires a **systematic and continuous search** for __manifestly illegal content__ online and its subsequent reporting so that it can take effect.

This *cannot* be left to the companies ("trust us!"), the **users** and the Police Internet Referal Units (IRUs) alone.

Third parties, e.g. civil society organisations, should be commissioned and financed to **systematically and continuously monitor** the platforms for **illegal extremist content**.

Status Quo: **About 450 policeofficers in IRUs monitor the web for 450 Million EU-Citizens.** *Manually.*

COUNTER
EXTREMISM
PROJECT

DSA consultations

CEP has already contributed to the first stage and will also publish a new policy paper where we will also focus, based on Hany Farid's research, the role of recommendation algorithms.

CEP POLICY PAPER

NetzDG 2.0

Recommendations for the amendment of the German Network Enforcement Act (NetzDG)

and

Investigation into the actual blocking and removal processes of YouTube, Facebook and Instagram

April 2020

© 2020 Counter Extremism Project Berlin | www.counterextremism.com | @FightExtremism

COUNTER EXTREMISM PROJECT

Policy Brief

Terrorist Content Online -
How to build comprehensible transparency for automated decision-making systems (ADM)

Alexander Ritzmann
Prof. Hany Farid

The dissemination of terrorist content is one of the most widespread and most dangerous forms of misuse of online services[1]. The current reporting mechanisms on preventing the dissemination of terrorist content, however, do not provide enough data or information to