



Issue 11 | August 2023

ONLINE SAFETY INDUSTRY UPDATE

For people who care about online safety from people who care about online safety

Today we are proud to launch our new podcast, [Radio SOSO - The Science of Safety Online](#). We talk about general issues and opportunities for the industry and are fortunate that many experienced online safety practitioners, with interesting stories and insights, have already had a chat with Martin.

Some of the other topics covered in this edition include:

- An approach that can unite free speech and online safety advocates
- Digital sweatshops, Red Teaming AI, and the potential for metaverse terror
- Regulatory, research and updates from around the globe



Uniting Online Safety and Free Speech

Arguments against new online safety interventions are often expressed as being in protection of free speech rights. But there are safety interventions that are effective at reducing harm – and protecting everybody's rights.

In this article we look at what a rights-based approach to online safety means, and what types of safety interventions best align with that approach.

It explains why human moderation teams within tech companies and helplines that act as brokers and trusted flaggers enable a genuine rights-based approach to safety online.

[Read the article](#)

Regulation Roundup

Updates from around the world



Australia

A revamped [classification scheme and new mis/disinfo regulations](#) are proposed

Also, eSafety issues a [position statement on generative AI](#).



USA

The [Kids Online Safety Act is amended](#), but [censorship and privacy concerns remain](#). The [Protecting Kids on Social Media Bill](#) is offered as an alternative, but [raises similar concerns](#).



Thailand

Thailand's digital minister [plans to ask a court to shut down Facebook](#) unless it takes action over scams.



Europe

The [Digital Services Act is now in effect](#) for the 19 online platforms designated very large.

Emerging Issues

Updates and information on emerging online safety challenges



The potential for metaverse terror

The Global Network on Extremism & Technology [examines some of the potential uses of the](#)



Exploring the Vulnerabilities of AI

"Red Team" events such as [Hack the Future](#) are [exploring and exposing](#) the



Digital sweatshops behind AI boom

AI is often thought of as human-free machine learning, the technology [actually relies on the](#)

[metaverse by terrorists](#) and suggests preemptive measures to minimise the potential risks. vulnerabilities of AI and could provide key information for regulation and protective measures.

[labor-intensive efforts](#) of a workforce spread across much of the Global South and often [subject to exploitation](#).

New Online Safety Exchange Podcast

Introducing Radio SoSo: The Science of Safety Online

Our new podcast is here! It is dedicated to general online safety issues and opportunities, and we are lucky to have already spoken to a stellar line-up of guests.

Subscribe at [Apple](#), [Google](#), or [Spotify](#).

Our first guest is Matt Soeth, Head of Trust and Safety, at Spectrum Labs. He talks about his background as an educator and how he first became aware of online safety - and the areas that that keep him engaged in this industry. [You can listen to our chat here.](#)



Industry Safety Updates

Safety updates and activities from the tech industry



Supporting targets of online hate

[ADL](#) Centre for Technology and Society [issues report cards for nine platforms](#) based on a list of 11 platform features recommended to support people targeted by hate.

Evolving T&S Vendor Ecosystem

A [growing ecosystem of specialist trust and safety organisations](#) are attracting venture capital and top talent - but the centralisation of key safety functions could also present risks.

A new approach to pornography?

It is a new [name and approach for the company behind Pornhub, YouPorn and Brazzers](#) as it promises to embrace innovation, diversity & inclusiveness, and trust and safety.

Research Updates



USENIX

Research [finds evidence of a significant gender gap](#) in terms of how men and woman seek out safety advice, and ultimately engage with security and privacy technology.



PNAS

[New research looks at the](#) dynamics of harmful content spread and finds the human moderator/ trusted flagger approach mandated by the The EU Digital Services Act remains effective.



Oxford Internet Institute

Large [independent scientific study finds no evidence](#) that Facebook's worldwide penetration is linked to widespread psychological harm.

Watch and listen

Podcasts and videos on relevant topics



Tech Policy Grind

David Ryan Polgar, Founder of All Tech is Human, [discusses what responsible tech means, and why](#) he thinks New York City is the center of the responsible tech community.



Tech Policy Press

Aditi Agrawal, Kamesh Shekar and Prateek Waghre [discuss India's Digital Personal Data Protection Bill](#) which joins a range of new tech policy laws and regulations in India.



Strangers on the Internet

Professor Bambauer [discusses dating app algorithms, safety, and possible tweaks to current tech](#) based on his expertise and personal experiences.

Expert commentary

[Justin Patchin of the Cyberbullying Research Centre looks at what the Surgeon General's Advisory on Social Media and Youth Mental Health actually says about the relationship between the two.](#)

[Yoel Roth, Tech Policy Fellow at UC Berkeley and former Head of Trust and Safety at Twitter, argues content moderation has a legalism problem](#)

[Jon Deedman from the Global Network on Extremism and Technology looks at how we can become more proactive in our response to the misuse of AI by increasing our education efforts.](#)

Contribute

The [Online Safety Exchange](#) (OSX) develops products and services to serve and support the online safety community – and works with international experts to achieve this. If you would like to contribute or feature something in the *Online Safety Industry Update*, please email martin@onlinesafetyexchange.org.

Join the OSX community

Online Safety Exchange

Auckland
New Zealand

You received this email because you signed up on our website or at one of our events.

[Unsubscribe](#)

