Teleperformance

T&S POLICY INSIGHTS NEWSLETTER

Issue No.2: Sept / Oct 2023

In this issue:

We highlight new regulatory updates across UK, Europe, U.S., and India. We discuss the latest trends in technology impacting T&S policies and enforcement practices and share some helpful resources for T&S professionals. Happy reading!

For more information contact:

Akash Pugalia, Global President, Media, Entertainment, Gaming and Trust & Safety and Farah Lalani, Global VP, Trust & Safety Policy.

Letter from the Editors

It is a busy time in the world of trust and safety with new regulations coming into force, technology advances challenging existing policies, and cost pressures continuing to put a strain on resources. Here are some of the trends we are noticing in T&S:

1 Push Back in U.S. Against New Online Safety Legislation

California Law related to Content Moderation Practices Challenged in Court

Elon Musk's X Corporation is suing California over its Social Media Transparency Law. This law requires social media companies with at least $100 million of gross annual revenue to issue semiannual reports that describe their content moderation practices and provide data on the numbers of objectionable posts and how they were addressed. The suit argues that this legislation violates the company's First Amendment rights because "it compels companies like X Corp. to engage in speech against their will" and "interferes with the constitutionally-protected editorial judgments" of the company. It will be interesting to see how this pans out given the implications for future legislation by states that aim to tackle 'lawful but awful' content.

Child Protection Law Blocked by Judge in California

A federal judge has granted Netchoice's request to block the California Age-Appropriate Design Code Act (CAADCA). This law requires special data safeguards for underage users online. The court's ruling stated the law likely violates the First Amendment. In August, an Arkansas court blocked an age verification law for users on social media, and a Texas court blocked a law requiring age verification for online pornography, saying that it would require invasive data collection and limit adults' access to constitutionally protected speech. Given legislation in Europe on age verification fast-progressing or already in place, it is unclear if and when U.S. legislation on this at the federal and state level will be aligned. (Link)

2 Attempts to Regulate AI – and Bridge the Gap to Regulatory Action

Voluntary Commitments on AI Safety Secured by White House

The White House has been securing voluntary commitments from leading AI companies. Some of the most relevant T&S commitments are around security testing of their AI systems before release, information sharing on AI risks, developing robust technical mechanisms to ensure that users know when content is AI-generated, such as a watermarking system, and publicly reporting their AI systems' capabilities, limitations, and areas of appropriate and inappropriate use; while these commitments are high-level, they seem to be appropriate in driving action at a time when so much safety legislation is getting tied up in U.S. courts.

AI Regulation Slow Moving as AI Development Speeds Up

While AI Regulation in China, Europe, Canada is emerging, enforcement is still years away. Nevertheless, the draft proposals of various AI bills provide interesting insights into the basis for the regulation. For example, Canada's Artificial Intelligence and Data Act would require that appropriate measures be put in place to identify, assess, and mitigate risks of harm or biased output prior to a high-impact system being made available for use; the obligations would be based on principles of Human Oversight & Monitoring, Transparency, Fairness & Equity, Safety, Accountability, and Validity & Robustness. Given the pace of AI development and the slowness of regulatory efforts, it remains to be seen how companies may proactively look to understand and mitigate risks through voluntary efforts. Meanhile companies are using voluntary frameworks such as the National Institute of Standards and Technology's AI Risk Management Framework.

3 New Research about the Potential Effectiveness of Regulation

New Study on Effectiveness of Content Moderation based on DSA Requirements

With the Digital Services Act (DSA) passed in November 2022 and now in force for Very Large Online Platforms (VLOPs) as of August 2023, it is heartening to read the findings of a new study published in Proceedings of the National Academy of Sciences, which aimed to model the effectiveness of this new regulation; it found that moderation efforts in line with the requirements of the DSA could effectively reduce harm, even on platforms with short content 'half-lives'.

Looking Deeper into UK's Online Safety Bill

With the Online Safety Bill now – and Ofcom taking the reigns to enforce the law as the regulator – the effects on user safety are ripe for further analysis. This interview between Susan Ness and Richard Allan delves eloquently into the nuances and the impacts of this landmark legislation, including the potential impact to small and medium size businesses, the impact to encrypted services, and of content legality decisions made by platforms.

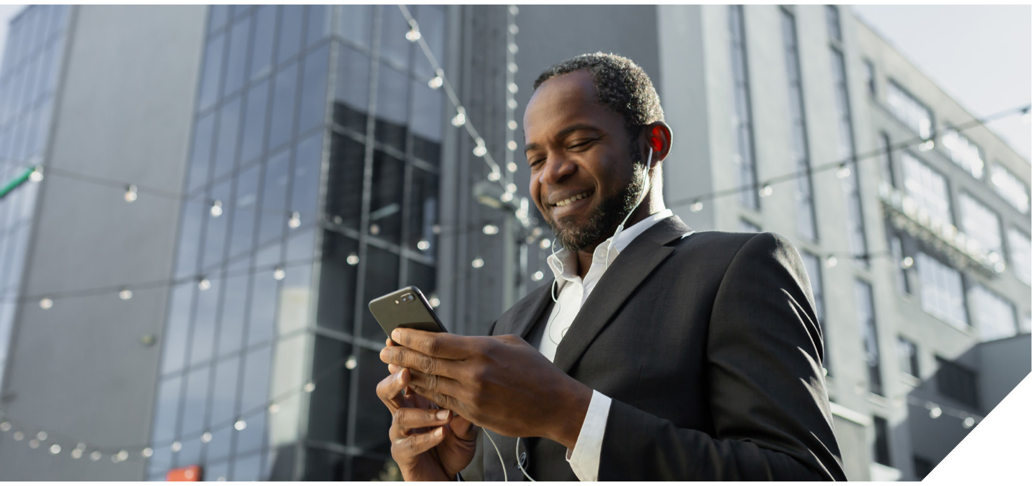
We hope you enjoy reading this issue of the newsletter and as always, feel free to reach out with any thoughts or questions!

Signed:

FARAH LALANI
Global VP, Trust & Safety Policy

AKASH PUGALIA
Global President, Media, Entertainment, Gaming and Trust & Safety

Platform Policies & Content Governance



Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms

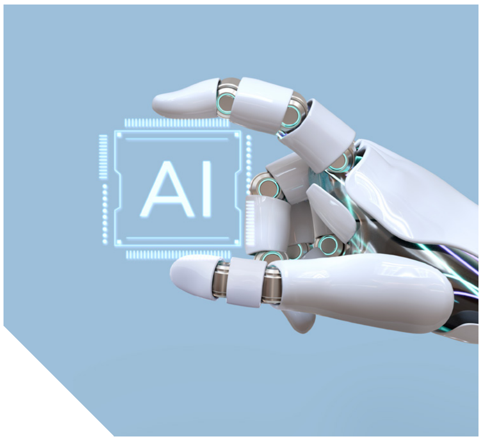
The World Economic Forum's Global Coalition for Digital Safety, with contribution from leaders at Teleperformance, has produced a Typology of Online Harms to create a shared vocabulary related to harms impacting people and society. It aims to define and standardize the typology of online harms in areas including threats to personal and community safety, harm to health and well-being, hate and discrimination, violation of dignity, invasion of privacy, and deception and manipulation. As online threats continue to evolve, the typology presents a vital step forward in achieving a collaborative, rights-respecting approach to digital safety, based on shared understanding and definitions.

Launch of Teleperformance Trust & Safety Advisory Council

On Monday July 10, 2023, Teleperformance launched the Trust & Safety Advisory Council – the first global initiative of its kind in the Digital Business Services Industry that brings together a multistakeholder group of leaders to foster a collective effort on internet safety. The Council brings together T&S leaders to share insights and spearhead relevant outputs related to new digital regulation, the impact of generative AI on harmful content detection, creation, and distribution, child safety, and content moderator wellness. A report highlighting the importance of content moderation for child safety had been produced with the International Centre for Missing and Exploited Children, one of the Council members, and more outputs of this kind are forthcoming.

AI-Generated Content Policies Take Shape in Advance of Big Election Year

Google will require election advertisers to add a 'prominent disclosure' to ads containing AI generated content that inauthentically depicts real or realistic-looking people or events to image, video, and audio content, across its platforms starting mid-November. Google highlighted that ads containing synthetic content altered or generated in such a way that is inconsequential to the claims made in the ad will be exempt from these disclosure requirements. Given, it seems, that the primary aim here to avoid showcasing content making it appear as if a person is saying or doing something they didn't do or generating a realistic portrayal of an event to depict something that didn't actually take place, it remains to be seen how this will impact election mis- and dis-information.



Regulatory Highlights

Digital Services Act In Force for Very Large Online Platforms and Search Engines

On April 25, 2023, the EU Commission adopted the first designation decisions under the Digital Services Act (DSA). This involved designating 19 Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) that reach a minimum of 45 million monthly active users. As of August 25, 2023, the DSA is now enforceable for these companies. Platforms must conduct risk assessments to understand potential risks in disseminating illegal content, negatively impacting rights, and intentional manipulation of users; while platforms remain non-liable for the content they host, they do have new requirements for content moderation systems, handling of notification of illicit content, prohibition of dark patterns and cooperation with law enforcement authorities.

Landmark Legislation Passed in UK

Much can and has been said about the UK's Online Safety Bill; it is a sweeping piece of legislation mandating social media platforms to remove illegal content quickly or prevent it from appearing in the first place, enforce age limits and age-checking measures, and provide parents and children with clear and accessible ways to report problems online when they do arise. These are just some of the many requirements placed on platforms which would have a legal responsibility to enforce the promises they make to users when they sign up, through terms and conditions. In light of recent investigations about animal torture content on social media platforms, the Bill also now covers animal cruelty. The fines and other penalties seem to be big enough deterrents to drive significant changes in the digital ecosystem; the guidance Ofcom provides to platforms will be critical in seeing how it is enforced.

Europe is leading the global race to regulate AI technology

The European Union (EU) is leading the way to take a definitive step towards regulating artificial intelligence. On June 14, 2023, the European Parliament approved the proposed rules, which seek to govern this rapidly evolving technology. A key aspect of the AI Act is a classification system that determines the level of risk an AI technology could pose within four risk tiers: unacceptable, high, limited and minimal. While high-risk AI systems may be permitted, developers must rigorously test these products, have proper documentation of data quality and an accountability framework that details how the system is subject to human oversight.

India Passes Digital Personal Data Protection Act

On August 9, the Rajya Sabha in India "unanimously" passed the Digital Personal Data Protection Bill (DPDP). One of the overarching goals of the legislation, by the government's own statement, is to strike a balance between protecting personal data and enabling the processing of such data for lawful purposes, so as to enable innovation and promote economic growth. An interesting implication of this law is that India will now require parent consent for the use of children's data including in social media.

Policy Insights

How should organizations and policy makers approach age verification?

As new regulations emerge to ensure child safety and protect children online, organizations must accurately determine the age of individuals. Policymakers, engaged in formulating and implementing online child safety regulations, face the challenge of determining the best approach to age verification. This paper aims to address this question by providing an overview of the escalating concerns among regulators regarding online child safety and examining international, national, and state legislation in this domain. Additionally, the paper explores the inherent tradeoffs associated with various age verification approaches, ultimately concluding with a range of options for regulators to consider when incorporating age verification into new legislation focused on online child safety.

Policymakers should enable greater adoption, as well as put in place guardrails to address the risks and ensure safe and responsible use of generative AI

Singapore's Infocomm Media Development Authority authored a report with Aicadium, a computer vision company, proposing ideas for senior leaders in government and businesses to build an ecosystem for the trusted and responsible adoption of generative AI. Key pillars of a safety framework it proposes follows the themes of Accountability, Data, Model Development & Deployment, Assurance & Evaluation, Safety & Alignment Research, and Generative AI for Public Good. The paper discusses how some of these may be practically achieved through watermarking and other such techniques. At a time when there is a lot of information about AI, it is nice to see a robust and easy to digest approach shared here.

Policy Tools & Resources

The Global Online Safety Regulators Network

The Global Online Safety Regulators Network is a global forum dedicated to supporting collaboration between online safety regulators and sending a strong message about shared commitment to online safety regulation. The Network was launched in November 2022, to share information, best practices, expertise and experience, to support harmonized and coordinated approaches to online safety issues. Current members include eSafety Commissioner – Australia, Coimisiún na Meán – Ireland, Film and Publication Board – South Africa, Korea Communications Standards Commission – Republic of Korea, Office of Communications (Ofcom) – United Kingdom, and Online Safety Commission – Fiji.

Global AI Legislation Tracker

The IAPP has published a tracker of legislative AI policy and related developments in a subset of jurisdictions around the world. It provides a brief commentary on the wider AI context in specific jurisdictions, and lists index rankings provided by Tortoise Media. It is the first index to benchmark nations on their levels of investment, innovation and implementation of AI.

Teleperformance