World Economic Forum Annual Meeting

Davos, Switzerland

Panel Discussion: Tackling Harm in the Digital Era

Suggested topics:

1. Safety by Design
2. Transparency and accountability, particularly in the context of Basic Online Safety Expectations
3. Digital Content Work Stream
4. Global Online Safety Regulators network.

Guiding Questions

1. The digital ecosystem is evolving quickly, with the advent of new technologies (e.g., immersive experiences, increased use of livestreaming), and an increasing level of regulation (e.g., EU DSA/ DMA, UK and Singapore Online Safety Bills). What actions did governments, platforms, and other stakeholders take, and what actions should be taken in the future to advance digital safety?
2. The growth of digital connectivity enables large populations of new users to access online services daily. What actions can the public and private sectors take to inform and protect users with different levels of digital maturity (e.g., digital native vs. new internet users)?
3. Platform governance has been at the forefront of the tech and media world recently, with important platforms shifting their approach toward content moderation and revisiting the balance of freedom of expression vs. safety. What actions should be taken to empower individuals to enjoy their human rights, promote healthy digital societies, and engender trust in an open, global internet? What principles should underpin decisions when it comes to balancing privacy and safety, knowing that sometimes enhancing privacy can pose risks to digital safety (e.g., end-to-end encryption)?
4. Digital safety challenges are exacerbated by the volume of activity across a broad, interconnected digital ecosystem. Harmful behavior is seldom confined to a single platform or geography, and tackling online harm requires diverse viewpoints in decision-making. How can governments, online service providers and civil society work toward more global cooperation to create a harmonized approach across jurisdictions? The Coalition members developed the Global Principles on Digital Safety, how can these principles drive multistakeholder alignment and enable positive behaviors and actions across the digital ecosystem?

Talking Points

Safety by Design

- We cannot hope to build a safer online world today, or into the future, if the fundamental building blocks of the Internet, Web 3.0 or the metaverse are not designed with safety at the

core and with the rights and wellbeing of individuals and society at the centre of such a design process.

- These imperatives need to be built in rather than bolted on – fundamentally, they need to be safe by design.

- Principles can provide really important scaffolding to guide the design, development and deployment of technology products and services and also enable a degree of flexibility so that companies are not shoe-horned into prescriptive feature mandates.

- eSafety went through the process of extensive consultation with industry through the development of our safety by design principles back in 2018. These principles covered three major areas, outlining service provider responsibility that development be human-centred, user empowerment and autonomy were really important elements and these were all underpinned by meaningful transparency and accountability.

- But, what we found is that there were a range of companies, from start-ups to enterprises that didn't understand what the risks and harms were, how their platforms might be exploited and what "good safety by design" looked like.

- So, in 2019 we took another 18 months to develop risk assessment tools, which we very much viewed as enabling companies to more purposely and consistently embed safety into their products and better comply with our Online Safety Act.

- These tools also surfaced up innovative safety best practices to demonstrate how forward leaning companies were addressing challenging safety issues so that these could be replicated across industries.

Online Safety Act and BOSE

- Our Online Safety Act was reformed and approved by Parliament in the middle of 2021 and next week will be the anniversary of the first year of implementation.

- One of our key systemic tools, the Basic Online Safety Expectations, built upon the key principles of safety by design and what the Government expected every online company serving Australians to provide in terms of fundamental online safety processes and practices.

- These Expectations are associated with extensive powers in the OSA to compel radical transparency from the technology companies. Transparency, that frankly the industry had not up until this point provided voluntarily.

- So, in December of last year, eSafety revealed the outcomes of the first use of these transparency tools. In a report that revealed the extent to which companies such as Apple, Microsoft, Meta, Snap and services or sites such as Skype, WhatsApp and Omegle were doing – or not doing – to tackle the hosting, distribution and live streaming of child sexual exploitation material and grooming on their services.

- These were actually carefully designed questions that many governments and child protection agencies had been seeking answers for years without meaningful answers.

- It also demonstrated that there was huge variation in terms of the time it took companies to address reports of child sexual exploitation – ranging from 4 minutes to 2 days or actually 19 days (if cases require 're-review'). Two companies didn't even have an in-app reporting function available.

- We also found that none of the video conferencing services were detecting livestreamed child sexual abuse, despite it being widely known that some of these services have been the primary vectors for this kind of abuse for some time.

- Unfortunately, it has taken legal powers to compel the type of radical transparency all of us in government need to see. Without that transparency, we cannot understand the true scale of the issues we're trying to tackle, nor can we hold these companies accountable.

- That is precisely why we are so pleased to see the European Commission coming on-board with the Digital Services Act and the UK's Online Safety Bill to join us in setting up clear legal frameworks and putting in place regulatory tools to mitigate online harms.

- We also recognise that fragmentation is a real issue so in November last year, Dame Melanie and I, and colleagues from Ireland and Fiji announced the formation of the Global Online Safety Regulators Network.

- The Network recognises that there will naturally be some differences in our regulatory schemes and approaches but there are areas where we can partner, learn from each other and ensure that we are coordinating our efforts wherever possible.

- We also hope that other nations and jurisdictions will see the benefits and what we are trying to do and be encouraged to join us.

WEF Digital Content Safety Principles and Workstreams

- Cooperation across sectors is also critical so it is great to see that the Digital Content Safety Principles just released by the World Economic Forum recognise that reducing online harms is fundamental to promoting human rights.

- Helpfully, these new principles build upon other established principles and frameworks including Safety by Design and the Voluntary Principles to Combat Child Sexual Abuse.

- But there is more important work to come. eSafety is partnering with Crisp Thinking to chair workstream number two, which will build on these principles and will strive to achieve a globally-recognised taxonomy of online harms – for now and into the future.

- We'll also look at how platforms and services tend to be exploited and misused and provide safety by design guidance and innovations that have been used across the tech industry to solve some of the most wicked problems.

Future Harms

- As well as understanding what technology companies are doing now to protect users online – and driving them to do better, where needed – we also need to be looking ahead, to anticipate the online safety challenges of the future.

- eSafety is continually scanning the horizon for these tech trends and challenges.

- Right now, we are very focused on the metaverse and what that means for the safety of both younger people and at-risk groups.
- While it's hard to know what form and shape the metaverse will eventually take, what we do know is that a range of tech-enabled, hyper-realistic and high-sensory experiences are coming together to form a totally new environment tipped by some to be worth $800 billion by 2024.

- Mark Zuckerberg appears to have bet the farm on this new reality, plunging billions into his Horizon Worlds platform, and even changed his company's name from Facebook to Meta to reflect his corporate intent.

- Essentially, immersive technologies will enable you to experience and interact with the digital world in three dimensions in a way that looks, sounds and feels almost real.

- Layer on haptics, which allow you to feel the internet – and what could possibly go wrong?!

- But the reality is that we can expect to see the same types of threats in the future metaverse as we see online today.

- And, while there will still be content children cannot unsee, it is the conduct – particularly in dark, private spaces – that will be cause for most concern.

- Harassment, abuse and cyberbullying are likely to have a far more visceral impact in more immersive environments, especially on children, as these digital experiences become more and more lifelike.

- We cannot afford to wait and see how things play out in these early versions before starting to prepare legal, regulatory and other policy frameworks.

- I believe we have a once-in-generation opportunity here to learn from our many Web 2.0 mistakes when building this bright and shiny new version, so that safety takes its rightful place as the third pillar of digital trust, alongside privacy and security.

- And this will require all of us to play our part, from users to governments, and of course the companies who must embrace a Safety by Design approach so future generations can use their products and services safely.

If asked about comments about freedom of speech needing to be recalibrated:

Avoid all the "re" words like "re-calibrate, "re-rebalance" and keep it to finding ways to balance competing rights online. And recommend talking about "protecting voices" rather than referencing "Freedom of speech". We want to show that this is something many countries around the world are thinking about so it's not just focused on us.

- I think there are a number of fundamental and often competing rights at play online and it's incumbent upon all of us to find ways to balance and protect them.

- It's fundamentally important we all have a voice, both in the real world and online, but it can't just be those with the loudest voices or those who hold the biggest megaphones who are heard at the expense of all others. After all, an equal vote and an equal voice is what true democracy is all about and I'd argue is its greatest gift.

- But we have seen time and time again that when online discourse veers into abuse, hatred, misogyny and violence, it can have a silencing effect on the person or group on the receiving end, which ultimately impinges on their fundamental right to have their voice heard, and to exist online free from such abuse.

- A parallel can be drawn to another fundamental right we all expect and deserve which is privacy online. But we also need to acknowledge the fact there are those who choose to misuse measures designed to protect privacy to avoid detection of their harm and sexual abuse of children.

- The sheer volume of child sexual exploitation and abuse material now circulating on the internet has reached epidemic proportions and I think we'd all agree that more needs to be done by governments, industry and at an individual level to tackle this issue.

- This might include looking at new technologies, and greater deployment of existing technologies, that preserve privacy while also protecting the rights of children to live with dignity and live free from sexual abuse.

- I think if this was an easy problem to solve, we would have done it by now, and I don't for a moment think that we in Australia have all the answers, but it's an extremely important conversation to be having, and conversations like these are often the first step to finding lasting solutions.

If needed: End-to-end Encryption

- While E2EE is of course very secure, like all technologies it can and has been misused to share illegal and harmful content like child sexual exploitation material and pro terror content.

- It's interesting that while Apple dropped its proposed client-side scanning for CSEM, it does scan children's encrypted messages on an opt-in basis to detect nudity. This demonstrates that potentially harmful material can be detected despite the use of E2EE.

- Dame Melanie will likely be able to talk in more detail about this, but the UK Government is currently running a Safety Tech Challenge Fund to encourage companies to come up with innovative ways in which sexually explicit images or videos of children can be detected and addressed within end-to-end encrypted environments, while ensuring user privacy is respected.

- So there is serious thought being put into how user privacy and children's safety can coexist in E2EE environments which is very encouraging. Again, I think it comes back to finding an appropriate balance of these digital human rights.

Background: Human rights in the digital world (from International team)

- Human rights are inherently interdependent, which means that in order to protect, respect and fulfill one right we have to work to fulfill all human rights. This includes human rights in relation to the digital environment.

- For example, if a person experiences discrimination on social media through bullying or hateful comments, it may negatively impact their emotional and mental health. This could make their participation in social media more challenging and they may be less likely to participate fully on social media in the long term. In this case not only are their rights to safety affected, but they are also not able to fully exercise their right to participate in online spaces.

- The ways that digital human rights intersect show that in addition to being interdependent, they are **indivisible**, meaning they can't be separated from each other, and they are **non-hierarchical**, meaning that all rights are equally important.

- The rights to privacy and freedom of expression have been given more attention in the early development of online spaces. However, the rapid expansion of access to digital spaces for people of all ages, backgrounds, locations and levels of experience with technologies – and the shift towards the digitisation of people's everyday activities – have made the right to safety increasingly important.

- Acknowledging safety as a right positions groups who are disproportionately impacted by harms online (such as women, children, racially marginalised groups, people with disability) as **rights holders** rather than 'vulnerable' or as defined by their exposure to risks or by problems they face on platforms. This positioning acknowledges the agency of representatives and organisations from these groups and their considerable work as advocates and change-makers in improving the safety and inclusivity of the online environment for everyone.