Journal & Proceedings of the Royal Society of New South Wales, vol. 157, part 2, 2024, pp. 266–284. ISSN 0035-9173/24/020266-19

How a single letter changed the world: W×3 – the World Wide Web (we weaved)

Julie Inman Grant

eSafety Commissioner, Australian Government[1]

Julie.InmanGrant@esafety.gov.au

Welcome from Her Excellency, The Honorable Margaret Beazley, Governor of NSW

As is our tradition at Government House, I welcome you in the language of the Gadigal: *Bujari gamarruwa Diyn Babana Gamarada Gadigal Ngura*. I pay my respects to their leaders, past, present, and emerging, as well as the Elders of all parts of our State from which you have travelled.

To say that computer technology has engulfed our world is like telling this audience in particular that Planet Earth isn't flat, although I can confirm that membership of the Flat Earth Society is free and joining up is simple. As of 2010, it had 189 members — at least, that's what I learned from Google.

I would also posit that even in this learned audience, some would be surprised at the extent to which computer and digital technology at the best, and often the very best, facilitates, but at the worst, controls, not only human interaction but many aspects of the everyday functioning of our society. This phenomenon is sometimes referred to as the Fourth Industrial Revolution. My own sense is that we are actually in an era of incomparable change.

Many benefits of this new era would have been unimaginable in 1822 when John Babbage, with funding from the Royal Society of London, designed what many consider the first computer,[2] and possibly not even in 1973 when the first commercial microcomputer, the Micral N, came onto the market.[3] Today, many of these benefits are aspects of everyday life to which we give little thought — digital phones, the internet, Google (that ready-at-hand encyclopedia of anything and everything), robotic surgery (which some of you here may have experienced), the ServiceNSW app and its QR code that got us through the doors of department stores during COVID and facilitates the renewal of our car registration, drivers' licences, and the payment of some of our taxes, internet banking, and, even more exotic, technology such as robotic security guards and robotic receptionists.

AI infiltration of the creative arts space has raised concerns of a different nature. As a teaser, I suggest you Google "Queensland Symphony Orchestra and AI" and read the associated articles to understand how AI can operate.

These and other challenges of technology, including its worst aspects, will undoubt-

[1] This paper is the transcript of a presentation in *Ideas@theHouse*, Government House, Sydney, 18 July 2024. See https://www.youtube.com/watch?v=nSFVrIugy3E

[2] https://science.howstuffworks.com/innovation/inventions/who-invented-the-computer.htm

[3] https://blog.wirelessmoves.com/2019/05/the-micral-n-and-others-the-micros-before-the-altair-8800.html

edly be the subject of interesting discussion tonight. By way of introduction, I would like to briefly refer to the reach of technology in areas that may not be so familiar to you. It's somewhat of a shopping list, and I won't dwell too much on the detail, but I found it interesting.

Recently, AI Steve, an AI chatbot, ran for Parliament in the constituency of Sussex, encompassing the towns of Brighton and Hove, in the recent UK elections under the banner of the SmarterUK Party[4] — I kid you not. AI Steve's human counterpart was Steve Endacott. The concept involved AI Steve having conversations with voters in Brighton and Hove, to ascertain the concerns of as many constituents as possible, and many more than traditional door-knocking would reach. If elected, the real Steve would represent AI Steve in Parliament. The real Steve explained that this campaign was a human-AI collaboration and described AI Steve as his co-pilot. AI Steve garnered 179 out of the 52,572 votes cast, so while not elected on this occasion, AI Steve might well say: "But, watch this space."

The tech company OpenAI, the maker of ChatGPT, recently discovered five operations based in Russia, China, Iran, and Israel using its technology to manipulate public opinion.[5] By this year's end, there will have been more than 64 elections worldwide: two recently concluded in France and Britain, and, of course, November 5 in the USA is just around the corner.

OpenAI instanced an example of an Israeli company called Stoic, which used OpenAI's technology to target social media accounts with pro-Israel content about the war in Gaza. Regardless of whether or not this involved fake news — and I have no way of knowing — it was at least a myopic representation of a very complex situation, demonstrating the danger of the misuse of technology, particularly to an illiterate and unsuspecting audience. The impact of the threat of AI on political influence is much more than a ripple on the surface of already troubled waters.

Law enforcement has become another area where technology is frequently used. In the US, one county has introduced a software product to draft police reports based on auto-transcribed audio from body-worn cameras during police operations. These reports, at the moment, are restricted to minor incidents and it is claimed that this will significantly increase police efficiency, given that individual police officers in the United States can sometimes spend up to 40% of their time writing up reports. The claim is that this "will prove to be one of the most impactful innovations of our time to help scale police work and revolutionize the way public safety operates."[6] We can hope, but we might have to wait and see on that one as well.

In New South Wales, child pornography is subject to a classification system from 1 to 5. Any form of child pornography is serious, but the level of depravity involved at the

[4] https://singularityhub.com/2024/06/13/say-hello-to-ai-steve-the-chatbot-running-for-uk-parliament/ and https://theconversation.com/britains-first-ai-politician-claims-he-will-bring-trust-back-to-politics-so-i-put-him-to-the-test-233403

[5] https://time.com/6983903/openai-foreign-influence-campaigns-artificial-intelligence/

[6] https://www.policemag.com/technology/news/15669250/axon-introduces-aipowered-automated-police-reporting-tool

higher end of the classification is beyond any reasonable sense of human comprehension. The impact on the police officers undertaking the classification is really shattering, and we've spoken to many of them. AI is now being used to classify the material in the first instance, subject to human verification. The reduction in the amount of material that has to be physically viewed by police officers is significant, with a corresponding benefit to the mental health of the officers undertaking this onerous but absolutely essential work. But it also needs to be said that technology has been the great enabler of the proliferation of child pornography in the first place.

Fortunately, for every technological advance, one can find the ingenuity of human or even anthropomorphic intervention. Australia now has 13 technology-detection dogs, not to track your Google usage but to detect various tech devices, including SIM cards and USBs. Recently, one dog found the phone of Samantha Murphy near a dam at Ballarat.[7]

For readers of yesterday's *Australian Financial Review* in hard copy, pages 11 and 12 had three articles on AI, one of which had the headline: "Artificial intelligence: Final nail in the coffin for the creative sector."[8]

I'm going to finish my shopping list here, and warmly welcome tonight's speaker, Julie Inman Grant. Following two decades working in senior public policy and safety roles in the tech industry at Microsoft, Twitter (now X), and Adobe, Julie was appointed in 2015 as Australia's eSafety Commissioner, a pioneering role leading the world's first government regulatory agency committed to keeping its citizens safer online.[9]

She was recently named one of Australia's most influential women by the *Australian Financial Review* and a leading Australian in foreign affairs by the *Sydney Morning Herald*. Earlier this year, she had the distinction of being named the "Australian censorship commissar"[10] — we quite liked that especially, because guess who called her that? Elon Musk, highlighting the irritation that some tech giants experience when governments seek to regulate content to protect the safety of their citizens.[11]

Julie, we are honoured to have you here. We look forward to hearing about your work in that world that the World Wide Web weaved.

Julie Inman Grant

Thank you very much, Your Excellency. It's wonderful to be with you here tonight. I would also like to pay my respects to the Gadigal people of the Eora nation and pay my respects to their Elders, past, present, and emerging. First, let me begin by thanking Her Excellency, The Honourable Margaret Beazley, Governor of New South

[7] https://ia.acs.org.au/article/2024/australias-technology-detection-dogs-revolutionise-policing.html

[8] https://www.afr.com/companies/media-and-marketing/ai-final-nail-in-coffin-of-australia-s-creative-sector-20240716-p5ju1u [Ed.]

[9] https://www.esafety.gov.au/about-us/who-we-are/about-the-commissioner

[10] https://www.bbc.com/news/world-australia-68878967 and https://www.internationalaffairs.org.au/australianoutlook/the-australian-esafety-commissioner-vs-x-testing-the-effectiveness-of-enforcement-powers-on-platforms/

[11] See (Ritchie 2024) for Elon Musk's behaviour towards the author. "Doxxing" is maliciously releasing a dossier of personal information of another. Brady J (2013), What is doxing? *Tech News*, April 2. [Ed.]

Wales, the Royal Society of New South Wales, and Susan Pond for inviting me to speak to you tonight. It's a great honour, not least because the calibre of the speakers who have stood here before me, an impressive crowd — accomplished Australians of all walks of life from politics, academia, science, and the arts. I'm humbled to join such an incredible group of people, and I look forward to our conversation further tonight.

How a single letter changed the world

This evening's theme was put to me by the Governor: "How a single letter changed the world."

When I first turned my mind to this intriguing idea, I must admit, I wasn't entirely sure whether it should be a letter of the alphabet I should be singling out, or a world-changing message, whether written by hand, keyboard, voice recognition or even AI. So tonight, I've decided to hedge my bets a little and highlight a pivotal example of each.

WWW

I'll start with my chosen letter of the alphabet, W. It represents one of the most important, world-changing inventions in human history — and, no, it's not the wheel. Well, actually I'm cheating a little here but, repeated three times, W becomes that famous acronym for the World Wide Web, a development which changed everything irrevocably. And it continues to exercise its transformative influence to this day, endlessly reinventing the way we live, work, play, learn and communicate.

This would be enough in its own right but I feel I can't talk about the Web without at least acknowledging another important letter, in this case an open one, dispatched, appropriately by email. While not as instantly transformative as the World Wide Web itself, this late-night missive generated quite some discussion and debate at the time, and it influenced the early philosophical underpinnings of internet governance, principles tech companies still hold onto doggedly to this day.

Before I delve further into that, however, I need to provide a little more context. There is no disputing that the World Wide Web of today is a truly remarkable thing which pervades almost every aspect of our lives, so much so that I think every one of us here would struggle to imagine our modern lives without it. But while it seems indispensable today, it really wasn't all that long ago that the internet was in its infancy and the world was still very much in an analogue state.

While computers have been networked in one form or another since the 1960s, it wasn't until the 1990s when affordable personal computers met the 56k modem that the internet really took off. These game-changing devices finally brought the internet to the masses, allowing people to slip on a pair of digital boardshorts and really start surfing the web.

While it was a time of great promise, many also felt over-regulation of this burgeoning industry would be an impediment to innovation and growth. Tech companies were "moving fast and breaking things" and wanted only one thing from governments — *to stay out of their way*. For the most part, governments around the world were happy to oblige.

I know this time well, having worked at tech policy "ground zero" in '96, as a young lobbyist in another significant W: Washington DC. Of course, being the '90s, I had big ideals, big shoulder pads, and even bigger hair, and eventually found myself working for an unassuming guy named Bill at a little software company called Microsoft.

It was Microsoft, AOL, AltaVista, Novell and Netscape that were the big end of town in the tech world at the time and — looking to the future through our rosy, techno-utopian glasses — we only had eyes for the *promise* this new online revolution would bring. Of course, this was Web 1.0 and social media wasn't really a thing yet — in fact, Mark Zuckerberg was just 12 years old and probably more concerned with building his latest Dungeons & Dragons adventure than the world's largest social network.

But with rapid advances in technology and computing power just around the corner, the internet we were preparing for all those years ago, would quickly outpace all of our imaginations.

John Perry Barlow writes a letter

Around the same time as I was walking the halls of the US Capitol, that other important letter I mentioned earlier was being written on a laptop at a raucous party in Davos in Switzerland during the 1996 World Economic Forum. This open letter was the now famous *Declaration of Independence of Cyberspace* penned by American poet, cyberlibertarian and occasional songwriter for cult US rock band "The Grateful Dead," John Perry Barlow.[12] In it, Barlow described a bold new vision of a completely free and open internet where people could reinvent themselves in this new virtual world with no government controls and no national boundaries. In short, governments should have no place in cyberspace.

While a revolutionary and somewhat controversial theory at the time, it was also an unashamed ode to a cyber utopia, focusing on the great promise this new world held for humanity, while giving little thought to how it might be potentially be misused to harm others, save for this one line. It said: *Where there are real conflicts, where there are wrongs, we will identify them and address them by our means.*

Now, I find this line compelling because I think it both mirrored and undoubtedly influenced how many of the tech leaders were approaching these issues in Washington DC at the time. Like Barlow, the industry I was then a part of also wanted the government out of cyberspace and pledged that if anything were to go wrong, they too would manage and moderate things themselves. Section 230 of the United States Communications Decency Act of 1996 helped enable and codify this, providing the industry, "intermediary immunity" from any bad acts or malicious content created by their users.[13]

Looking back, it was probably a little naïve of Barlow and the politicians in Washington to believe that such a new and untested industry could be left to its own devices and be trusted to self-regulate. But even then, the compelling industry arguments that regulation would put the brakes on innovation, economic growth and US tech hegemony were hard to resist.

Needless to say, Barlow's late-night email was one of the earliest examples of viral online manifesto and is still widely shared today. I was lucky enough to have the opportunity to have a spirited debate

[12] https://en.wikisource.org/wiki/A_Declaration_of_the_Independence_of_Cyberspace

[13] See https://www.pbs.org/newshour/politics/what-you-should-know-about-section-230-the-rule-that-shaped-todays-internet for a 2023 discussion of this section. See also Smith and Van Alstyne (2021) [Ed.]

around these concepts with Barlow at an expat Thanksgiving dinner party in London in 1997. But that is a story for another evening, in less polite company!

But this brings me to another triple-W representing one of the key questions being asked about today's online world: *what … went … wrong?*

What went wrong?

The views Barlow expressed in his declaration are emblematic of much of the public discourse we've been having around the internet for decades and continue to have today. Time and time again, governments have let their eagerness to reap the promised benefits of shiny new technologies blind them to any of the potential pitfalls and we are still yet to totally reconcile this today. Whether it's launching a new search engine to bring information to the masses, a social media platform to give a voice to those who previously had none, or disrupting an established industry to deliver better value and choice to the consumer, it is more often than not born out of a sincere belief that whatever the creation might be, it will make society better.

But I think today one could equally argue that many of these benefits have also come at a great cost, not just to individuals, but to society more generally, especially when the race to be first to deliver the latest product to market continues to outweigh the responsibility to ensure it is safe. The current industry scramble to be first to bring Artificial Intelligence to the masses is just the latest example of this flawed tech ethos. If we don't learn from our past mistakes and put in the safety guard rails now, I fear the damage that could be wrought on society by these immensely powerful programs could be irreversible.

To that end, I do think that 2023 reached a tipping point when generative AI came so quickly into the mainstream. Governments finally took notice of not only the long-term existential threat to humanity promised by the arrival of Artificial General Intelligence, or AGI, but of the immediate online harms we are already starting to see play out today.

2023 was also the year governments coalesced around AI Safety — and a year when many other countries joined Australia by passing key online safety legislation and setting up independent online safety regulators — specifically in the UK, Ireland, and across the European Union through the Digital Services Act.[14]

eSafety also set up and chaired the Global Online Safety Regulators Network,[15] a formal body that would help ensure cross-border collaboration and information sharing in online safety regulation. After all, the Internet is global, whilst laws are local, and the only way we are going to encourage a very entrenched and powerful industry to start minimising harms at-scale, is by working together. And the only way we are going to create a safer, more civil online world is to ensure that the fundamental building blocks are "safer by design."

In some ways, the internet serves as a mirror reflecting societal ills, but there's no doubt it can serve to amplify and even weaponise them.

[14] https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

[15] https://www.esafety.gov.au/about-us/who-we-are/international-engagement/the-global-online-safety-regulators-network

The age of digital divisiveness

In this age of unbridled digital connection, modern society has probably never been more polarised than it is today, with public discourse and respectful debate replaced by increased intolerance and division, as algorithmic echo chambers reinforce and often provide false legitimacy to more and more extreme views. We are truly entering the age of "digital divisiveness."

Sadly, much of this outrage-driven engagement has been the "by-design" guiding principle, to keep platforms sticky; and with business models driven by paid influencers, surveillance advertising and subscriptions, those who are most outspoken or controversial profit the most. As a result, democratic ideals and institutions have never been under such grave threat. The January 6th riots at the US Capitol building after Donald Trump's election defeat in 2020 gave us a terrifying glimpse of the power and influence the internet holds over us all and just how fragile and delicate the fabric holding democracy and society together can be.[16]

In some ways, January 6th should not have been a surprise. During his presidency, Donald Trump not only savagely abused foes online with impunity but was identified as a major "superspreader" of mis- and disinformation. However, his online audience was so sizeable and his content went so viral, that none of the major platforms suspended him for repeated policy violations, making excuses for vaguely worded "exceptions" for public figures.

A 2021 analysis of 120,000 posts and tweets on Facebook and Twitter (now X) found that there was just a small collection of 12 individuals and their organisations responsible for the vast majority of anti-vaxx misinformation circulating on the global Internet during lockdown.[17] Whilst Donald Trump could not claim the dubious distinction of being named one of the "Disinformation Dozen," another US presidential candidate, Robert Kennedy Jr., *is* amongst them. Again, Instagram, Facebook and Twitter failed to enforce persistent violations of their own house rules until the incitement of online violence so clearly spilled into real world harm that they *had* to act.

Even this past weekend, where we saw political leaders of all stripes condemning the political violence brought so clearly into focus by the attempted assassination of Donald Trump, the unhinged fringe and conspiracy theorists spun mistruths more quickly than the government and journalists could uncover and report the facts. This came from both the Left and the Right. And the hashtags "Civil War" and "Fake Assassination" trended across social media in a matter of moments. And so, even though today we are blessed with instant access to almost limitless information at our fingertips, we've never been more distrustful of it. Indeed, the online world is starting to resemble an ever-expanding desert of misinformation with fewer oases of truth in which we can find refuge.

Unfortunately, humans are going to have their critical reasoning skills even more challenged with the accessibility, ease of use and

[16] The riots across the UK after the killings in Southport, with Elon Musk's encouragement, are a further example of the power of misinformation on social media to create mayhem in mature societies. See https://www.bbc.com/news/articles/cd1e8d7llg90 [Ed.]

[17] https://counterhate.com/wp-content/uploads/2022/05/210324-The-Disinformation-Dozen.pdf

low cost of powerful AI applications that allow anyone with a smartphone to create photo-realistic deepfakes of people saying and doing things they didn't really say or do. Deepfakes have now become "cheap-fakes" that are easy to create and difficult to discern by even the trained eye. They cost virtually nothing for the perpetrator to make but exact a devastating, lingering and incalculable cost to the victim-survivors.

Without content provenance requirements and an acceleration of accurate and rapid deepfake detection tools, there is a real risk that a deepfake could become the viral narrative before the real truth is revealed. Like many have said before, unbridled AI could not only upend free and fair elections and continue tearing at the fabric of society but it could also ruin lives.

Bullying

The Web, and the technology that underpins it, has enabled new kinds of abuse that didn't exist before its creation. While children have always faced the threat of bullying, today's internet has made it so much more pervasive and invasive, heightening its impact on a child. Where once a child would find respite when they walked out of those school gates, thanks to smart phones, social media and messaging apps, bullying now follows children 24/7 into their homes and bedrooms.

eSafety administers the only youth cyberbullying complaints scheme in the world and we have a 90% success rate in getting this harmful content taken down when the platforms fail to act. But we continue to witness concerning trends. In the year to May 2024, complaints to us about cyberbullying of children were 311% higher than the same period four years ago. Most of this continues to be peer-to-peer but we're also seeing children as young as nine years old reporting to us. This is part of a larger trend post-lockdown where parents were much more permissive with children and their technology use.

Complaints about image-based abuse — or the non-consensual sharing of intimate images — were 242% higher, driven primarily by sexual extortion reports. As a result of this trend, the demographics of intimate-imagery abuse have markedly changed. At the inception of our image-based abuse scheme in 2018, 70% of image-based abuse targeted women and girls — often as a consequence of relationship retribution.

Today, overseas criminal organisations are targeting young men between the ages of 16 and 24 for large sums of money once they have tricked or coerced them into creating and sharing sexual imagery of themselves. In some very tragic cases, these young victims have taken their own lives rather than face the shame of seeking help from friends or family. Meanwhile, other forms of image-based abuse, including abuse related to family and domestic violence issues such as coercive control and the incidence of "deepfake image-based abuse" have also surged.

Reports of illegal and restricted content were 111% higher. We investigated some 33,000 URLs over that 12-month period, 85% of which concerned suspected child sexual abuse material. Last year, we also reported that one in eight complaints about this material now relate to self-produced child sexual abuse coerced remotely by a predator online, and often captured in the child's bedroom or bathroom of the family home — literally under their parents' noses.

Add to this overlay of clearly definable harms that eSafety is seeking to tackle with tools under the Online Safety Act,[18] the next generation of harms we have not fully reckoned with are already on their way. Whether it's the prospect of "augmented telepathy" or the robbing of "cognitive liberty" through invasive technologies like neural implants, we need to start preparing now.

Age restrictions on internet users

To be sure, the latest wave of techno-panic being exported from the US is already washing over our shores warning us that screen time is re-wiring kids' brains and social media is the sole cause of teens' anxiety and depression. This was given more heft when the US Surgeon General recently suggested public health warnings should be applied to social media, much like those emblazoned on cigarette packets to warn of the potential health risks posed.[19]

Some argue the tech industry is already acting like Big Tobacco and should therefore be treated as such, as the industry is accused of ignoring compelling research that shows the damage its platforms pose to children so tech firms can protect their bottom lines. This debate might even see Australia swinging the pendulum much more into the interventionist camp of online safety regulation, with a media-fuelled push banning children under the age of 16 from joining social media.

While this will ultimately be a policy question for Government,[20] I think a much deeper debate needs to be had around what we mean by social media — because kids aren't posting to Facebook like they were a decade ago, but are using a range of different platforms and instant messaging services. We also need to think long and hard about what the unintended consequences might be of pushing kids into darker recesses of the Internet. I'm also concerned the pursuit of forbidden online fruit will deter help-seeking and confiding in parents when things do go wrong online.

I can tell you that the evidence base is thin and the research very mixed on all of these questions. But the teens who are the most marginalised and vulnerable today are the most likely to be impacted by such a ban. Our research clearly shows that LGBTIQ+ teens, First Nations youth, and those with a disability feel more comfortable and themselves online than they do in the real world and depend on technology to connect, explore and find their tribe.

To be honest, as Australia's regulator in this area, I'm struggling to get my head around the "how" in terms of successfully implementing such a policy. You see, I have been working on age assurance in one shape or form since 2008. In 2023, eSafety delivered an age-verification roadmap to government. A trial is now underway and I've also given industry six months to create meaningful codes to prevent children from accessing pornography and other high-impact content up and down the tech stack: on devices, through app stores, in search engines, and on social media. But until these fundamental age-assurance technologies and associated safeguards are in place,

[18] https://www.legislation.gov.au/C2021A00076/latest/text

[19] https://www.theguardian.com/us-news/article/2024/jun/17/surgeon-general-warning-social-media

[20] The Australian government plans to ban under-16s from social media platforms. Nov 6, 2024. https://www.abc.net.au/news/2024-11-08/how-the-age-minimum-for-social-media-will-work/104571790 [Ed.]

implementation and enforcement of such a ban will be virtually impossible.[21]

Or water safety awareness

The best analogy I can give to you in terms of a recommended approach starts with another W: water safety awareness. Australia has built and excelled in water safety over the past several decades, often in response to tragic circumstances — and this has important lessons for successful approaches towards online safety. In short, we fence pools and we back these rules with enforcement; but we do not try to fence the entire ocean. On our beautiful beaches, we protect our citizens from a young age teaching them to swim, to stay between the flags and recognise rips to avoid danger. We also put in shark nets to protect from predators that might be lurking just beneath the water. But no matter how well we prepare our kids, we still keep a close eye on them in the water because we well understand the dangers.

We should be using the same philosophy online. We need to ensure our kids have the confidence and digital literacy they need to navigate the online currents safely, while teaching them how to spot the dangerous algorithmic rips and lurking predators. Parents should act as the digital lifeguards, keeping a close eye on our kids while still allowing them to dip their toes in the ocean. And our extensive parent content on https://www.esafety.gov.au/[22] helps empower parents to do just that!

The world's first online safety regulator

This now leads us to *why* eSafety was stood up as the world's first online safety regulator in 2015 and how we approach this work. As you can imagine, there was no playbook available for online safety regulation, so we've had to fill in the pages as we've gone along. In some cases, we've had to write the playbook.

Prevention

Our approach and model is multi-pronged beginning with *Prevention*. Through our research, education, and awareness-raising programs, we strive to prevent online harms from happening in the first place. These start in the early years, as 91% of kids have access to a digital device by the age of 4, all the way to over-65s through our Be Connected program.[23] But meaningful and lasting societal change takes time and until then, Australians suffering harm will continue to reach out to us for help.

Protection

This is where our *Protective* powers come in. Under Australia's Online Safety Act, eSafety operates several world-first schemes to protect Australians online. I touched on these issues earlier but these include our child cyberbullying scheme, our serious adult cyber-abuse scheme, and our image-based abuse scheme. Through these complaints-based regulatory schemes, we support individuals in the grip of personal online crises by compelling social media platforms and websites to take down abusive and

[21] Australian parliamentary inquiry stops short of backing social media ban for under-16s; see https://www.aph.gov.au/Parliamentary_Business/Tabled_Documents/8267 [Ed.]

[22] https://www.esafety.gov.au/parents

[23] https://beconnected.esafety.gov.au

harmful content. We also have remedial powers that target both perpetrators and platforms. In my estimation, this is one of the most unique and important functions we have — we serve as a safety net and can remediate harm quickly when the platforms fail to act.

Proactive and systemic change

Our daily engagement with Australians gives us insights into concerning online trends and also provides us ample evidence of where the companies are failing at a systems and process level. I'll touch on our systemic powers around transparency, codes and standards in a bit, but I wanted to touch upon our third pillar, "*Proactive and systemic change*," which is not enshrined in our legislation but is critical to being an anticipatory regulator and in shifting the burden of safety back onto the platforms themselves.

Safety by Design

Key to this is the Safety by Design initiative[24] eSafety launched in 2018, something we did *with* rather than *to* the industry. We have to be realistic that we will never arrest or regulate our way out of online harms, so we thought it reasonable for companies to assess risks upfront, understand how the harms manifest against their users, and incorporate safety into every aspect of how they design, develop, and deploy their products and services.

Victoria's 1970 seat-belt legislation

The best way to explain this approach, I have found, is through real-world analogies. I'm going to take you back even further — more than a half century — to remind you of a young American lawyer named Ralph Nader, who began publicly questioning car makers' accountability for traffic fatalities; critically examining the intersection between vehicle design and a lack of embedded safety features like the humble seatbelt. His ideas culminated in a book called *Unsafe at Any Speed* — and this led to a new era of automobile regulation, still guided today by international standards.

Predictably, at the time, the auto industry vehemently pushed back. They didn't want to invest in seatbelts or any other safety features, believing the ongoing costs would be prohibitive and would stifle both profits and automotive innovation. Of course, today we know the seatbelt alone has saved millions of lives and we take for granted that every car is now brimming with life-saving technologies that are built in — like airbags and anti-lock brakes. In fact, today, car manufacturers differentiate themselves in the marketplace based on their safety ratings and consumers take note. Proof that safety sells!

Yet, despite the appeal to consumers and to serious harms reduction, it took legislative bodies around the world to compel the embedding of seatbelts into cars for these to become standard fare.[25] There are certainly a number of comparisons we can draw here

[24] https://www.esafety.gov.au/industry/safety-by-design

[25] From December 22, 1970, the state of Victoria became the first jurisdiction in the world to make wearing of seatbelts mandatory while travelling in a car. Front seatbelts had been mandatory in all cars sold in Australia since January 1, 1969. https://www.carsales.com.au/editorial/details/buckled-to-history-21137/ [Ed.]

with the current state of technology regulation.

I believe we are fast approaching the tech industry's "seatbelt moment" — and not a moment too soon, as a new industry race has begun to bring AI to the masses and to be the first to colonise the "metaverse." But history shows us that it's unlikely companies will make this important transition to safety-first technology voluntarily — shiny new gadgets that please consumers and securing first-mover market share is pleasing to shareholders, so will almost always win out. We also see piecemeal retrofitting of safety features announced by press release — with no data from platforms on take-up or efficacy.

The simple truth of the matter is, if your platform has 50 different parental controls that parents have to toggle on, then you haven't used safety by design as a fundamental development principle. Likely the only way you are going to make your service truly safe is by totally re-engineering it.

Australia's Online Safety Act (2021)

For this reason, Australia's Online Safety Act (2021) includes important systemic powers aimed at applying significant pressure on the industry to bring about meaningful change — and that involves looking under the hood.

The first of these powers is Basic Online Safety Expectations,[26] which includes wide-ranging transparency powers that compel companies to answer key questions about how they are living up to these expectations and tackling a range of online harms. Under these powers, we've sent 19 notices covering 30 major services including Apple, Google, Meta, Microsoft, TikTok and X Corp, asking questions about how they are tackling a range of harms, including child sexual abuse, terror and violent extremism, sexual extortion, online hate, and harmful algorithms.[27]

Just getting these companies to finally lift the lid and reveal some of their inner workings is a *win* in itself. For decades governments around the world have been asking these same questions with little success. We are now sharing this data with our international partners to help them better understand what *is* and *is not* being done. It continues to be my belief that sunlight acts as the best disinfectant.

Through the operation of these powers, we've seen positive online safety outcomes. Unsurprisingly, some of these changes have come about through the "naming and shaming" aspect of these powers — as it is generally reputation and revenue impacts that are more likely to move companies toward the light, rather than regulation alone.

We understand that this will be a constant battle. Just as governments are achieving greater levels of transparency about what is actually happening within the metaphorical bowels of these platforms, we are, in fact, seeing movement from the major industry players to become more opaque.

Placing Application Programming Interfaces (APIs) out of reach, threatening litigation, personally targeting regulators and justices, acquiring and then deprecating potent social media monitoring tools like

[26] https://www.esafety.gov.au/industry/basic-online-safety-expectations [Ed.]

[27] See Root and Ashford (2024) [Ed.]

CrowdTangle[28] are just the opening gambits. Governments need to continue pushing harder and staying a step ahead.

Another way we are trying to get and stay ahead and encourage active systemic change is through the implementation of mandatory codes and standards to tackle online child sexual exploitation and pro-terror material.

As an example of another very significant win, six world-first codes are now in operation across eight sectors of the technology ecosystem. While two of these codes produced by industry did not meet appropriate safety community safeguards required under the Act, I was able to use my powers to write the rules for them moving to industry standards. These two standards cover broad groupings of tech services like cloud-based file and photo-storage services, gaming and dating sites, and messaging services.

I cannot impress upon you how pivotal these two standards are in protecting children, as we know that cloud-based storage services and encrypted messaging are used widely by paedophiles and terrorists to store and distribute this incredibly damaging content. The standards have been registered with the Parliament and — following the usual due process — should also come into force around Christmas.

But, of course, the more things change, the more they stay the same, and dogged industry resistance to any regulation remains an ongoing challenge. For years we've seen some big tech businesses throwing their weight around Down Under, challenging Australia's approach on important issues like harmful online content, child protection and even payment for news.

The stark headlines are nothing new, nor are they unique to our shores and we should not be swayed by them. For example, some high-profile industry members affected by our standards were so worried about the global implications of what we were asking them to do that they mounted what can only be described as a good old-fashioned fear campaign to sow public and policymaker doubt.

While solely focused on forcing the industry to do more to prevent their services being misused by paedophiles to harm children, one leading company (named after a fruit) even went so far as describing eSafety's draft standards as a "Dystopian Dragnet" which would inevitably lead us down a slippery slope of mass government surveillance of ordinary, law-abiding citizens.[29]

The pushback from encrypted services was just as fierce, despite eSafety being explicit that we do not expect industry to break or weaken end-to-end encryption. But we were equally explicit that it was no longer good enough for encrypted services to throw their collective hands in the air and do nothing either. This is a form of wilful blindness and serves as another example of the industry once again prioritising the absolute privacy of adults to undertake any act "in the dark" without considering the dignity and commensurate rights of children to live free from online violence and abuse.

We can and will weather the pushback from one of the most wealthy, stealthy and powerful industries in modern history — what other choice do we have but

[28] Meta's CrowdTangle was no longer available after August 14, 2024. [Ed.]

[29] Apple warns that scanning encrypted photos leads to a "dystopian dragnet," *The Stack*, Sept. 4, 2023. https://www.thestack.technology/apple-photo-scanning-csam-dystopian-dragnet/ [Ed.]

to continue standing up and pushing the regulatory barrow against this long-term "technological exceptionalism"?

After the Bondi Junction stabbings

Part of the job of a regulator is to test the efficacy of the powers we have today. Those powers were certainly put to their first real public test in April 2024. You may recall in the wake of the Bondi Junction tragedy there was a graphic high-impact video of an attempted murder of a bishop delivering a livestreamed sermon at his church in Wakeley. The attack was declared a terrorist incident by the NSW Police Commissioner, committed by an allegedly radicalised teenager: more than 52,000 potential terrorist and violent content images were later found on his phone.

No surprises there that exposure to harmful terrorist content will desensitise, normalise and even radicalise impressionable young minds.

Following the Christchurch atrocity of 2019,[30] the Government proscribed a very specific role for eSafety when a potential livestreamed attack like this occurs. We conduct rapid-fire, online investigations to determine how far and how quickly this high-impact and gratuitous violence is spreading to determine whether it should be deemed an "online crisis event." We then notify the social media companies of the content and assess what steps they are taking to stem its viral spread and protect innocent eyes from stumbling across something they would never be able to unsee. Our sole goal and focus is to prevent extremely violent content from going viral, and in the case of the Wakeley attack, potentially inciting further violence and inflicting more harm on the Australian community.

And so, on 16 April, eSafety issued formal removal notices to Meta and X Corp — requiring both companies to take all reasonable steps to ensure the removal of this extreme violent video content. The removal notice identified specific URLs where the video material was located on both of these services.

Their responses could not have been more different. Meta complied within the hour, following up consistently with updates on the steps they were taking to ensure this abhorrent content was not re-loaded and re-shared.

But as we all saw, X Corp not only refused to remove the content but vigorously defended their right to keep hosting the video of a brutal attempted murder on its platform in the Federal Court, despite the fact that the video more than likely violated their own terms of service.

And this brings me to one of the most disappointing Ws of them all.

Weaponisation

While many companies acknowledge their societal responsibilities in a moment like this, some companies choose *Weaponisation* of their platforms for profit and *warfare* with regulators, or more accurately "all out lawfare" against those who try to bring more safety and civility to their platforms, rather than set any global precedent that compliance might create.

X Corp is sadly such a company and in this case, its barristers deftly avoided addressing the harmful nature of the content, rather directing focus instead on the more abstract

[30] Dobbins (2019) [Ed.]

concepts like the comity of nations; freedom of speech under the US First Amendment; and even whether a decision might subject the Federal Court to international ridicule. As a result, we saw very clearly that this was a particular legal battle we were not going to win, so I chose a strategic withdrawal so that my powers could be tested before the Administrative Appeals Tribunal (AAT). Fear not, the legal battles with X remain ongoing …

We are involved in five more separate actions with X Corp, either in the Federal Court or the AAT, including a key Federal Court battle around X Corp's non-compliance with a transparency notice asking questions about what the company is doing to combat child sexual abuse material and failure to pay the infringement notice.

But herein lies the rub. These companies have almost unlimited funds to tie regulators up in multiple, lengthy and costly litigations. It's a strategy I like to call "death by a thousand courts" and one they have repeated in other parts of the world. In fact, Elon Musk just threatened to take the European Commission to a "very public battle in court" for preliminary findings that X Corp violated the Digital Services Act for a range of deceptive design features.

But this process has unearthed a bigger, more fundamental question which all tech regulators and countries outside the US will need to grapple with: if technology companies like X Corp are not answerable to the laws of the sovereign nations in which they operate and extract revenue, then to whom are they ultimately answerable?

An answer to this pivotal question is well beyond the boundaries of one litigation or the jurisdiction of a single court and will have sweeping implications for all digital platform regulators. We need to ensure these decisions are made in the right forum with a broad range of decision makers.

A way forward?

So, how do we find a *way* forward? One thing that's clear is that the ability of all sovereign nations to protect their citizens from harmful online content and conduct needs to be part of a global conversation. Perhaps it is a convention, or a treaty, or the equivalent of a "Bretton Woods"[31] for cross-jurisdictional regulation of online harms.

As mentioned, eSafety has aligned with other global safety regulators through a formal organisation and recently signed a regulatory MOU with the European Commission, a formidable bloc with plenty of regulatory heft behind them.[32]

But of course, the issue of regulation and sovereignty will likely continue to be challenging until the most significant jurisdiction where most global tech giants are headquartered — the United States — makes serious attempts to pass safety legislation and hold them accountable.

Things did seem to be moving in a positive trajectory with an emotionally-charged US Senate Judiciary Committee hearing in January 2024, but further progress has stalled. Of course, the outcomes of the upcoming US election can and will have reverberations for technology regulation, regardless of the outcome. How can it not?

[31] The 1944 Bretton Woods Conference led to the regulation of the international monetary and financial order after the conclusion of World War II. [Ed.]

[32] https://www.esafety.gov.au/newsroom/media-releases/esafety-partners-with-the-european-commission-to-support-enforcement-of-online-safety-regulations

But, importantly, while these court battles have been going on, the Australian Government has also brought forward the Online Safety Act Review[33] and I have no doubt the recent experience with X Corp. and any weaknesses exposed in our legislation during this case will be examined closely to ensure future legislation is fit for purpose. Some options on the table for consideration include bringing our powers into line with those of the UK and European Union, with powers to fine companies up to 10% of annual operating revenue as well as significant business disruption powers.

In any meaningful plan to find a way forward, I think it's hugely important there is a greater degree of coherence and coordination between global regulators. While our regulatory systems will never be identical, it's important they are aligned to increase our collective effectiveness in regulating these powerful, US-domiciled companies.

I know that was a lot and I want to thank you so much for listening.

Recently, 2021 Nobel Peace Prize winner Maria Ressa[34] labelled today's tech CEOs as the "world's largest dictators." Maria is someone whom I deeply respect — and has tangled with political dictators throughout her journalistic career — so certainly knows tyranny when she sees it. But I think she makes a valuable point about a seemingly untouchable class of powerful and politically potent tech billionaires and their followers.

When you really think about it, today's tech leaders wield almost unlimited power, not just over a single country, but over a captive global populace. They also don't just hold themselves above the law but seem to exist completely beyond it, backed by almost inexhaustible financial resources.

Up until his death in 2018, Barlow steadfastly stood by his Declaration of Independence of Cyberspace, but I do wonder, if he was still alive today, and witnessed so much power and influence in the hands of so few in his cherished egalitarian cyberspace, whether he would start to have some second thoughts.[35]

The World Wide Web has undoubtedly changed the world, but we can't just continue to celebrate the good while turning our backs and ignoring the bad. This is a world we wove — or rather, a web spawned in the great US of A — and there haven't been meaningful guardrails to prevent these harms.

Think of how far the online world has come in the past 30 years from the dial-up days, and yet Section 230 of the Communications Decency Act has not been touched since 1996.[36] But there is hope. I believe we are making quiet but solid progress and — with more countries and jurisdictions creating their own online safety regulators coming online — we are no longer alone in this fight. eSafety was once the sole voice in the wild calling for change — we are now hearing these calls from others around the world echoing like a steady drumbeat. Safety by Design is taking hold globally; governments are asking global tech companies to

[33] https://www.infrastructure.gov.au/sites/default/files/documents/online-safety-act-2021-review-issues-paper-26-april-2024.pdf [Ed.]

[34] https://www.sipa.columbia.edu/news/nobel-laureate-maria-ressa-join-sipa-faculty [Ed.]

[35] See https://en.wikipedia.org/wiki/John_Perry_Barlow [Ed.]

[36] See Solomon (2024) for recent cases, research, and actions in the US. [Ed.]

prioritise safety and well-being ahead of tech profits and shiny new gadgets.

Can we do it?

And this brings me to the final and most important W of all — *we*. It really is only by working together, and showing regulatory courage and cooperation across borders and jurisdictions that we can hope to change how the technology game is being played.

Ever the optimist, I remain eternally hopeful that, if we do this, we can wrest back control of this incredible world-changing invention and have the safer and more inclusive world wide web we all want, in line with what its creators always envisaged. Thank you.

Questions and Answers

Susan Pond: Thank you, Julie. Thank you very much. That was incredible. I don't know where the 45 or 50 minutes went, but they've flown by. I only saw a few people's heads bobbing. My name is Susan Pond, and I'm honoured to be the President of the Royal Society of NSW. In partnership with Her Excellency in Government House, we have held *Ideas@theHouse* ten times now, and it hasn't disappointed. We've had amazing presentations, and yours is right up there in the pantheon. I'm going to open it up to the audience.

Q1: Thank you, Julie, for that absolutely fantastic presentation. I'd like to thank you very much for the work you and your part of the government do to protect our children. I was very interested in your analogy with the car industry. Having worked for a large company for 27 years myself, I saw the internet go from a nice utopian place to one with a very dark side. When cars were introduced, the government imposed safety measures by insisting manufacturers install seat belts. Is there something similar we could do for the internet? We all access it through physical devices. Could we ask manufacturers to install safeguards, akin to installing seat belts? I know it's not a perfect solution, and there would be challenges, but I'm wondering if there are ways we can use our sovereign power to build in some solutions to address this problem?

JIG: Well, the codes I mentioned actually divide the technology industry into eight different sectors, and phase one of the codes took almost three years to develop. That's why I gave the industry six months. We put together a paper for them, basically saying we need choke points and safeguards throughout the stack on devices, through app stores, and search engines. That's one way. With the Online Safety Act review, I suspect there will be an approach around a duty of care, with Safety by Design being a fundamental element of that. You're absolutely right: when we import Tesla cars to this country, we expect them to meet Australian standards. Otherwise, they don't come in. But this technological exceptionalism and jurisdictional arbitrage we're seeing will continue to challenge us. Have you seen Andrew Forrest's litigation against Meta for scams? He's taking it to California to meet them where they are. One technicality in the judicial review that X Corp is challenging us on involves Nevada incorporation law and whether it carries duties for foreign jurisdictions. A lot of people in government thought the jurisdiction question was answered, but I don't think it has been.

Q2: Thank you for your presentation. We really admire and respect the work you do for us. We can't deny that the rapid rise of technological advancements comes with many benefits, but, as you mentioned, there's also misuse of technology. How do you think young people can help create a safer online environment?

JIG: I think role-modelling good digital behaviour is the first step. I want to mention that key members of my staff are here, and I'm just one person; I couldn't do it without them. One of my special guests here is Professor Joanna Weaver, who developed the Technology Policy Design Centre. People often ask if I always wanted to be the eSafety Commissioner. I didn't; there was no internet, and I had no idea I would be living in Australia. But I think you make your own luck, and I've been lucky to be in the right place at the right time, just when technology policy was happening in Washington DC. Before there was an internet, I was working for my hometown Congressman on social issues. One day, he said, "We've got this small company in our electorate called Microsoft. Can you work on technology issues too?" So, I started working at the intersection of social justice, policy, and technology before there was an internet. That was lucky, and then I created a path to pursue an interest. Many people are turned off by technology, but technology policy is for those of us who are right-brained and want to bring social justice to technology. Of course, legal degrees help as well.

SP: Your passion is infectious, and the work you're doing is extremely important. Thank you for sharing it with us tonight. We're all in a world where it's difficult to determine the difference between right and wrong in the metaverse — the tensions between transparency and censorship. We applaud you for the work you're doing. It's not easy, as you've demonstrated. We're all with you in the broader society as we see the struggles in our children and grandchildren and the attempts to regulate in a world where technology far outpaces policy. Can you tell me your greatest nightmare? Is it that the technology is outpacing policy and the gap is probably getting wider?

JIG: Actually, it isn't. My greatest nightmare is that someone will take their own life after they've come to us, whether due to cyberbullying or experiencing sexual extortion, and we couldn't help them.

SP: Yes, I'm sure you have many nightmares. Would you please join me in thanking Australia's eSafety Commissioner, Julie Inman Grant.

References

The Disinformation Dozen: why platforms must act on twelve leading online anti-vaxxers. Center for Countering Digital Hate, 2021. https://counterhate.com/research/the-disinformation-dozen/

Dobbins J (2019) The Christchurch massacre was another internet-enabled atrocity, 20 March. https://www.rand.org/pubs/commentary/2019/03/the-christchurch-massacre-was-another-internet-enabled.html

Nader R (1965) *Unsafe at Any Speed: The Designed-In Dangers of the American Automobile*, Grossman.

Ritchie H (2024) "These aren't just words:" The woman threatened for taking X to court, *BBC News*, September 6. https://www.bbc.com/news/articles/cx2ymd32g2eo

Root J and Ashford G (2024) Inside a high-stakes fight to limit social media's hold on children, *New York Times*, 29 March. https://www.nytimes.com/2024/03/29/nyregion/social-media-algorithms-children.html

Smith MD and Van Alstyne MW (2021) It's time to update Section 230, *Harvard Business Review*, August 12. https://hbr.org/2021/08/its-time-to-update-section-230

Solomon A (2024) Has social media fuelled a teen-suicide crisis? *The New Yorker*, October 7, pp. 26–37. https://www.newyorker.com/magazine/2024/10/07/social-media-mental-health-suicide-crisis-teens