



Diligent design in an age of digital abrasion

ThinkPlace's submission to the
AHRC's Human Rights and
Technology consultation

ThinkPlace creates impact through human-centred innovation and design.

Our mission is to contribute to the United Nations' Sustainable Development Goals, and in doing so, create vibrant societies, strong economies, sustainable environments and trusted institutions.

Emerging technologies, and the new models they enable, offer the potential for great benefit and great harm to everyone in our society, from the most resilient to the most vulnerable.

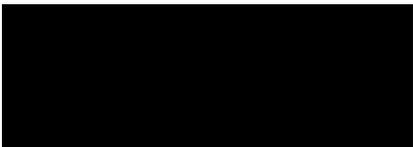
ThinkPlace cares deeply about the impact of digital technology on society.

We recognise that, as a community, we are in a transition period where the pace of innovation is so great, not even activism and social license, much less legislation and regulation, can comfortably keep pace.

In this context, we feel that the AHRC's human rights and technology consultation has the opportunity to provoke important thinking on an issue that will affect all of us, and perhaps start a nuanced conversation that is much needed and very timely in Australia.

In this context, we are pleased to share our thinking on important aspects of the digital transformation of Australian society. They have been built up through our work with many of the most vulnerable populations in Australia and in the world more broadly; and through our collaborations some of the most sophisticated public good institutions and organisations in Australia and internationally.

We are happy to stay involved and be an ongoing contributor, as the AHRC intent is strongly aligned with our public good mission.



Darren Menachemson
Global Partner
Global Chief Digital Officer
Canberra, Australia



On behalf of the ThinkPlace Australia Executive

The need for Diligent Design	5
The road to digital abrasion: a predictive analytics case study.....	6
The road to digital abrasion: a technology-facilitated abuse case..... study	7
Digital abrasion and the ethical challenge of new ICTs.....	8
Characteristics that make digital abrasion hard to manage.....	8
Mitigating and attenuating the risks of abrasive digital innovations.....	9
Contact.....	10

The need for Diligent Design

In the coming decade, we confidently predict that every sector in Australia will have easy access to cheap, powerful digital technologies, made widely available as cloud-based services.

By the late 2020s, integrating AIs into business models will become trivial.

Adopting simple, commodity-level intelligences will become as simple as adding a shopping cart to your cloud-hosted business website is today. Adopting more sophisticated, tailored intelligences will become a minor technical challenge, equivalent in complexity to exploiting an open-data API. Creating widespread or micro-targeted communications is now as simple as signing up to the right social media engine.

In this imminent world of democratised machine learning, predictive analytics, and mass/micro-communications, we expect to see powerful benefits for, and harms to, both individuals and to society as a collective. There is a significant danger that the pace of innovation will be much greater than can be matched by legislation, regulatory design or even activism, and meaningful responses will lag behind the risks they seek to mitigate.

This paper suggests that to attenuate such harms and reap such benefits, society needs to set not just guardrails defining what is acceptable and unacceptable, but also embrace an expectation of Diligent Design. We define Diligent Design as a design process that:

- 1. Meaningfully engages with those that may be affected** (users, their families, their community) by an innovation – particularly, people experiencing vulnerability – to understand the potential for the innovation to cause harm. This could include user research, literature reviews or direct consultation.
- 2. Meaningfully evaluates innovations prior to their release**, to see if there are likely to be unintended consequences that need to be attenuated or designed out.
- 3. Meaningfully assesses if the innovation would enable human rights violations** - breaches of privacy or unlawful **discrimination**.
- 4. Establishes the presence or lack of social license**, actively seeing to understand whether the proposed innovation passes the “sniff test” and ensuring that social, financial and personal costs to the individual or community are transparent, minimised and balanced by genuine benefit.
- 5. Designs-in an authentic informed consent process** that ensures progressive informed consent and avoids unnecessary “all-or-nothing” situations that compel uninformed or reluctant consent.

The road to digital abrasion: a predictive analytics case study

In 2014, RMIT and Monash University collaborated on a study that studied how social media posting habits of individuals with bipolar disorder differed from neurotypical individuals. The intent was to understand if algorithmic analysis of posting behaviour (e.g. regularity of posting, time of posting) could consistently identify if a consenting person with the condition was having a manic episode.

The individual would then be alerted. Alerts would then escalate to someone they nominated (e.g. a family member or a clinician).

In 2018, a project in National Tsing Hua University in Taiwan has gone a step further, using a machine learning algorithm to scan twitter feeds for indicators of bipolar disorder (e.g. word use, lingual energy levels, posting frequency). Rather than seeking consent, they look for people who may not know they have the disorder, so that they can then (hypothetically) be identified and encouraged to seek treatment.

Last month, the Canadian Government's Public Health Agency (PHAC) released a tender entitled "Artificial Intelligence (AI) pilot project for surveillance of suicide-related behaviours using social media", stipulating that the project should not seek to interact with the 'studied' and should be a non-intrusive public health surveillance tool.

And today, you can go onto Kaggle and download an open source kernel to perform analysis of correlations between mood and twitter posting behaviour.

These milestones create a pathway to a possible future where:

- ✓ The cost to society of undiagnosed or untreated mental illness is reduced
- ✓ The personal tangible and intangible burdens of unmanaged mental illness are potentially alleviated
- ✓ Policy-making and programmatic interventions are improved
- ✓ People with mental health challenges, and their carers, can subscribe to helpful digital services that can improve the coordination and timing of care
- ✗ People have intrusive and disturbing experiences without their consent where they learn about a mental health condition without a healthcare provider being present
- ✗ Marketers take advantage of manic states to encourage gambling or irresponsible purchasing behaviour
- ✗ Employers, insurers, credit providers and discriminate against people with mental illnesses by filtering them out of opportunities
- ✗ Peers and friends covertly assess people with mental illness using consumer-targeted tools, breaching health privacy and creating isolation and loneliness
- ✗ Society loses trust in technology and social license for positive/trustworthy initiatives is lost

The road to digital abrasion: a technology-facilitated abuse case study.

Our recent work for the Department of Social Services in consulting on the Fourth National Plan to Reduce Violence Against Women and their Children has identified a significant increase in technology-facilitated abuse.

Technology-facilitated abuse can include texting, sexting, revenge porn, digital stalking, digital tracking/monitoring, and the malevolent use of social media). It is typically perpetrated by men against women. Technology has given a new set of tools to these perpetrators of domestic abuse, and they using them to create serious harm.

The continuing rise of technology-facilitated abuse means that both the nature of, and the evidence of, violence is changing. There is no standard response to this in a variety of federal and jurisdictional systems (e.g. justice systems). The ubiquitous nature of social media and the more ambiguous justice and regulatory environment means that women (including older Australians), children and young people are increasingly exposed to technology-facilitated forms of violence.

Technology has created a significant gap in the system that needs to be addressed.

At one level, this must be undertaken through awareness raising, education and the development of substantial and integrated policy and service responses to these forms of violence. Australia has seen some policy and support initiatives in this area and needs to maintain and increase the strength and comprehensiveness of its response.

At another level, the way technology is used to facilitate violence is evolving as quickly as the technology itself so it very hard for the system (services/police/courts) to keep abreast of how it is being used and how to respond to it. In some cases, a critical intervention point must rest with the providers of those technologies and their application of Diligent Design to design down if not design out the capacity for these types of abuse (or at least design in facilitated responses that expedite sanction towards the perpetrator).

Digital abrasion and the ethical challenge of new ICTs

Australia's government and community institutions, as well as its business community, will need a strong competency in navigating new dimensions of the ethical challenge we term 'digital abrasion'.

Digital abrasion is the tension that emerges when what governments, businesses, NGOs and individuals *can* do with digital and data comes into conflict with what the community deems as being *acceptable*. From a human rights perspective, it is the threat of causing harmful breaches of privacy, risks to personal safety or enabling unjust discrimination.

Rushworth Kidder, the eminent American ethicist, distinguished moral dilemmas from ethical dilemmas. He considered moral dilemmas as 'right vs wrong' questions, and ethical dilemmas as being 'right vs right' questions, where one has to decide between important principles that have come into conflict with each other. Digital abrasion comes from both moral and ethical dilemmas.

Sometimes, digital abrasion can be a morals-driven issue, that is reasonably easy to diagnose. To reference Kidder, we may ask "Are there obvious 'wrongs' associated with an innovation that means it would fail the Mom Test (would you be ashamed to tell your mother about it?), or the Stench Test (does it intuitively feel corrupt or malevolent?)?"

Other times, and perhaps more challengingly, digital abrasion can be an ethics-driven issue where there are conflicting 'rights' associated with an innovation (e.g. balancing the risk of causing harm vs the risk of foregoing opportunities to reduce suffering; or benefiting the collective good at a cost to the individual good).

Characteristics that make digital abrasion hard to manage

Digital abrasion is difficult to navigate, because it often shows a number of troubling traits that confound traditional design processes. These include:

- **Novelty** - They have no or few precedents
- **Impact** – They create the risk of substantial, scaled impact on the community
- **Momentum** - "Do nothing" is not an option because of expectations and events beyond one's control.
- **Heterogeneity** - Community expectation and social licence is not homogenous, but instead fragmented across every grade of the spectrum of opinion
- **Complexity** – The potential benefits and harms are difficult to explain and difficult to map comprehensively
- **Sensory lag** - Results of an intervention will not be predictable nor emerge quickly
- **Accumulation** - Each failure will make successive attempts more and more politically and socially untenable, delaying the desired outcome

Mitigating and attenuating the risks of abrasive digital innovations

There is no clear answer as to how the issue of digital abrasion will be solved; it is a classic wicked problem, and like all wicked problems, will need significant co-design, experimentation and a multi-faceted systems approach.

However, based on what we hear from the community when we engage with them as ethnographers and anthropologists, and what we see happening in governments and other institutions we work with on equally complex challenges, we feel that any solution will likely need to include:

- **The adoption of an ethics framework** to guide the design of potentially abrasive innovations, which includes a strong statement on the expectation of **Design Diligence** (as described above on page 4).
- The development of a **principles-driven, enforceable regulatory model** that takes a **differentiated treatment approach towards ICT providers**, recognising different compliance behaviours (e.g. responding differently to innovators showing voluntary design diligence, an unintentional lack of design diligence, an intent to cause deliberate harm, or an intent to create serious/systematic harm).
- **A well-coordinated regulatory environment**, recognising that many government agencies (e.g. the Australian Competition and Consumer Commission, the Australian Media and Communications Authority, the Digital Transformation Agency, the Australian Federal Police, the Attorney-Generals Department, the Department of Social Services, the Commonwealth Ombudsman, the Fair Work Ombudsman, the Department of Foreign Affairs and Trade, the Office of the Australian Information Commissioner and of Australian Human Rights Commission) will all own and control part of the policy and regulatory solution. Creating and co-designing a joined-up system that integrates rather than fragments government efforts will be essential in dealing with the risk effectively.
- **Building a strong digital ethics authorising environment, capability and culture in the Australian Public Service.** This will be needed to ensure that the APS is in a position to develop effective policy, regulatory and service interventions that protect the Australian community from careless or deliberate harm, and encourage and incentivise a Diligent Design approach both from the APS itself, and from the business and NGO community more broadly.

We would be happy to continue the discussion. To connect, please reach out to:

Darren Menachemson
Global Partner

