

CHRISTCHURCH CALL

TO ELIMINATE TERRORIST
& VIOLENT EXTREMIST
CONTENT ONLINE

Regulatory Implications for the Public Service

Presentation to **ipanz**, 16 October 2019,
Te Puni Kōkiri Wellington



Peter A. Thompson
Media/Communication Studies
Victoria University of Wellington



Aims of presentation-

- Evaluate how far the Christchurch Call summit in Paris provides a workable policy framework to a) curtail online terror/ extremism and b) respond to the wider regulatory/policy issues raised by social media and digital intermediaries.
- Interrogate the motives behind Mark Zuckerberg's *welcoming of regulatory intervention* and the new willingness of the tech companies to engage with governments.
- Argue that the Christchurch Call is a positive starting point but extremist/ terrorist content online is just the *tip of the iceberg*- There is a need for a **wider regulatory framework** for regulating social media & digital intermediaries.
- Consider some of the implications of this for the ongoing work of government departments and the need for a) a **domestic policy response** alongside supranational efforts, and b) **enhanced collaboration** across the public policy sector.



The Context-

- On 15 March 2019, a terrorist used **Facebook live-stream** to broadcast the mass murder of 51 at two Christchurch mosques.
- The initial 17 minute live-stream was watched by 200 people, none of whom reported it.
- It was only after *29 minutes* that the video was reported to Facebook -by which time it had been viewed by 4000 people.
- Over the next 24 hours, Facebook removed **1.5 million** uploads of the video, 1.2 million of which were blocked prior to upload- but 300,000 could not be identified by algorithms.
- The video was widely uploaded to other media including YouTube and extremist sites based overseas like *8chan* and *Kiwi Farms*.
- Some excerpts of the video were still being found circulating on **YouTube, Facebook and Instagram** over a month afterwards- one *Youtube* video had generated >720,000 views.



```
import socket, sys, os
print "[Attacking]" + sys.argv[1]
print "injecting" + sys.argv[2]
def attack():
    #pid = os.fork()
    s = socket.socket(socket.AF_INET,
    s.connect((sys.argv[1], 8011)
    print ">>> on"
```

The Christchurch Call- Paris summit, 15 May 2019

- Initiated and chaired by Jacinda Ardern and France's Emmanuel Macron with a stated aim to “***eliminate terrorist and violent extremist content online***”
- Brought together 17 governments + the European Commission and 8 major social media and digital intermediaries including **Facebook, Google, YouTube and Twitter.**
- Premised on the recognition of a need for a ***supra-national, multi-lateral*** response to the issues, rather than unilateral, domestic state regulation.
- The summit did produce a non-binding **pledge document** outlining principles and responsibilities of governments and online service providers
- Although aspirational, evidencing more ‘what’ outcomes than ‘how’ mechanisms, the summit was arguably a success in providing a platform for future multilateral discussions.



- Counter the drivers of terrorism and violent extremism by strengthening the resilience and inclusiveness of our societies to enable them to resist terrorist and violent extremist ideologies;
- Ensure effective enforcement of applicable laws that prohibit the production or dissemination of terrorist and violent extremist content;
- Encourage media outlets to apply ethical standards when depicting terrorist events online;
- Support frameworks, such as industry standards, to ensure that reporting on terrorist attacks does not amplify terrorist and violent extremist content;
- Consider appropriate action to prevent the use of online services to disseminate terrorist and violent extremist content.



CHRISTCHURCH CALL

Online Service Providers:

- Take transparent, specific measures seeking to prevent the upload of terrorist and violent extremist content and to prevent its dissemination on social media and similar content-sharing services, including its immediate and permanent removal, without prejudice to law enforcement and user appeals requirements;
- Provide greater transparency in the setting of community standards or terms of service;
- Enforce those community standards or terms of service in a manner consistent with human rights and fundamental freedoms;
- Implement immediate, effective measures to mitigate the specific risk that terrorist and violent extremist content is disseminated through livestreaming;
- Implement regular and transparent public reporting;
- Review the operation of algorithms and other processes that may drive users towards and/or amplify terrorist and violent extremist content to better understand possible intervention points;
- Work together to ensure cross-industry efforts are coordinated and robust;

facebook

Google

You Tube

twitter

amazon

Microsoft

Qwant

dailymotion

CHRISTCHURCH CALL

Both Government and Online service providers

- Work with civil society to promote community-led efforts to counter violent extremism in all its forms
- Develop effective interventions, based on trusted information sharing about the effects of algorithmic and other processes, to redirect users from terrorist and violent extremist content
- Accelerate research into and development of technical solutions to prevent the upload of and to detect and immediately remove terrorist and violent extremist content online
- Support research and academic efforts to better understand, prevent and counter terrorist and violent extremist content online;
- Ensure appropriate cooperation with and among law enforcement agencies for the purposes of investigating and prosecuting illegal online activity in regard to detected and/or removed terrorist and violent extremist content;
- Support smaller platforms as they build capacity to remove terrorist and violent extremist content;
- Collaborate, and support partner countries, in the development and implementation of best practice in preventing the dissemination of terrorist and violent extremist content online'
- Develop processes allowing governments and online service providers to respond rapidly, effectively and in a coordinated manner to the dissemination of terrorist or violent extremist content following a terrorist event.
- Respect, and for Governments protect, human rights, including by avoiding directly or indirectly contributing to adverse human rights impacts through business activities and addressing such impacts where they occur;



CHRISTCHURCH CALL

Update: United Nations, New York, 24th September

- New **crisis response protocol**, (for governments and tech companies) to coordinate and to manage the online impacts of terrorist/ violent extremist incidents.
- Establish a **Christchurch Call Advisory Network** to advise on the implementation of responses.
- **31 new countries** signed up to the Call (total now 48) while the Council of Europe and UNESCO also joined the European Commission as international bodies.
- **Global Internet Forum to Counter Terrorism (GIFCT)** is to become a formal body to lead the tech sector's response. **3 Pillars:**
 - **Prevention** e.g. education and redirection.
 - **Response** e.g. increased cooperation with stakeholders & sharing information with law enforcement.
 - **Learning** e.g. support research into digital platforms and evaluate best practices for multi-stakeholder cooperation.



- GIFCT has also been developing a **Content Incident Protocol**, a '**Counterspeech Toolkit**', and sharing **hashing algorithms** to detect extremist content with smaller tech companies.
- **Google & YouTube** have also been working to strengthen responses to extremist content and have recently tightened restrictions on live streaming and hate speech-
- e.g. In the second quarter of 2019, 80% of the 9m videos YouTube deleted were removed before viewing. It also modified its algorithms to favour more 'authoritative' sources in search results.

For its part, **Facebook** has committed to a 9 point response:

- **Terms of Use** updated to prohibit the distribution of terrorist and violent extremist content.
- **Improved User Reporting** of Terrorist and Violent Extremist Content –*clearer categories* for users to flag extremist content.
- **Enhancing Technology** for algorithmic detection of extremist content including investing in **digital fingerprinting** and AI.



GIFCT
Global Internet Forum
to Counter Terrorism



- **Livestreaming** –more vetting measures and active moderation of streamed content .
- **Transparency Reports** published on detection/removal of extremist content.
- **Shared Technology Development**- sharing data sets and open source detection tools.
- **Crisis Protocols**- establishment of incident management teams.
- **Education**- including user advice about sharing/reporting content.
- **Combatting Hate and Bigotry**- supporting research on impact of online hate on offline discrimination and violence.
- In September 2019 Facebook also started redirecting NZ users viewing extremist content to websites helping people exit hate-communities.



- Although the Christchurch Call summit understandably focused on ***curbing online terrorist/extremist content***, it comes in the context of a *wider trajectory toward domestic and regional state interventions* in the activities of social media and digital intermediaries.
- This stems from a growing acknowledgement of the **political, economic and civic harms** attributable to the operations of social media and digital intermediaries including-
 - **Privacy/surveillance/security issues related to mass harvesting of personal data;**
 - **Control over audience content discovery and facilitation of filter-bubbles/proliferation of fake news/ misinformation;**
 - **Enabling interference with electoral processes and undermining democracy.**
 - **Impact of digital intermediaries on traditional media value chains/business models and systemic avoidance of taxes.**



Cambridge
Analytica

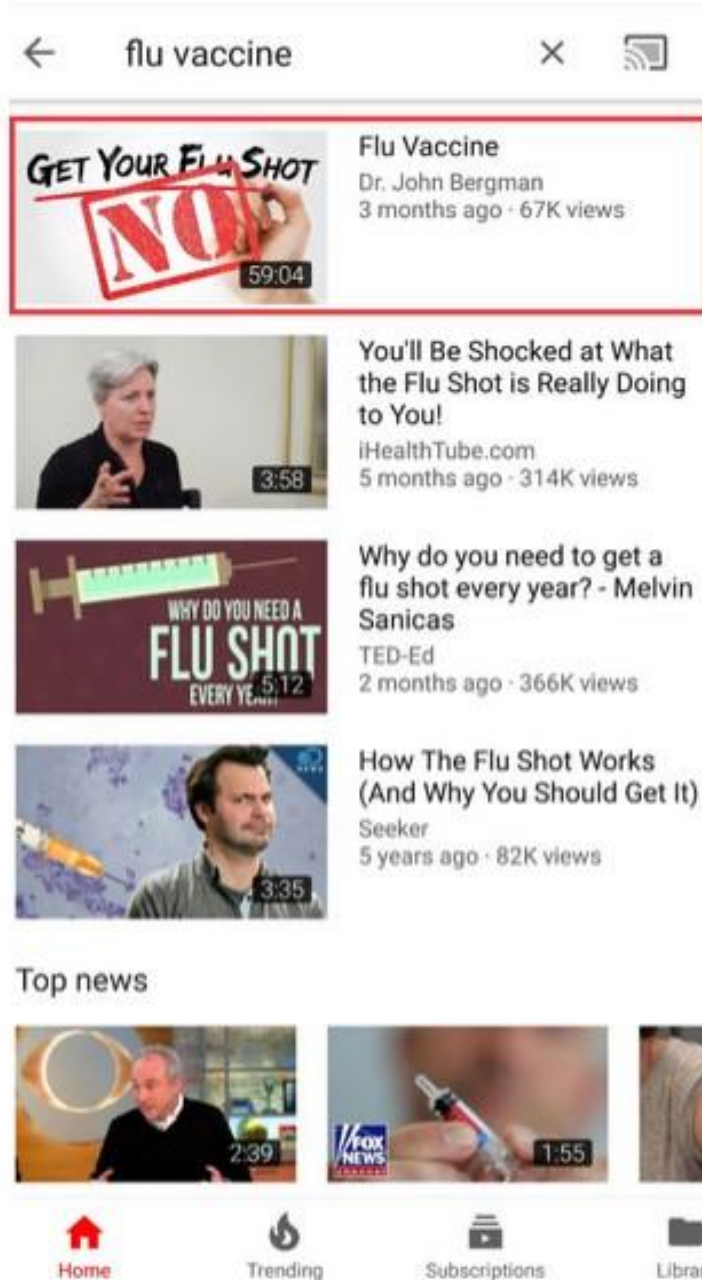
- Mark Zuckerberg has called for *“a more active role for governments and regulators,”* saying that, *“Regulation could set baselines for what's prohibited and require companies to build systems for keeping harmful content to a bare minimum.”*
- But recent investigations of social media content moderation practices revealed serious **conflicts of interest** in moderating extremist content that also generated lucrative online traffic:
- UK Channel 4's **Dispatches** uncovered Facebook's 'shielded review' policy which referred far-right content breaching community standards up the corporate ladder if it generated *high traffic levels*.
- Vice's **Motherboard** found that while Facebook blocked **white supremacist** content, it allowed references to white **nationalism/separatism** to remain (now discontinued post-Christchurch).
- Motherboard also uncovered evidence that Twitter was reluctant to block some extreme right wing comments because of the potential controversy if *elected Republican politicians* were affected!



- A 2018 [Bellingcat](#) online micro-ethnography analysed how 75 fascist activists became radicalized or '**redpilled**'. This suggested that more *moderate* right-wing content encouraged a '**pipeline**' toward *increasingly extremist* material. Of the 75 activists, 6 cited **Infowars**, 10 cited **4chan**, and 15 cited **YouTube**.
- Another 2018 study by [Vox-Pol](#) examined the 28000 Twitter followers of 41 self-identified **Alt-right Tweeters**. One finding was a prevalence of **YouTube** and **Facebook** (also Instagram and WordPress) as sources being shared among the group.
- These have recently been corroborated by an international study led by the [University of Minas Gerais](#) which examined >330,000 **YouTube** videos on 360 channels categorized as *Intellectual Dark Web, Alt-lite and Alt-right*.
- Analysis of 79 million comments and user activity showed that there was a distinct **pattern of migration from moderate to extremist content**. YouTube *algorithm recommendations* appeared to be a contributory factor driving this.



Searching for 'flu vaccine' brings up these video recommendations:



If a user selects the first video, the recommended videos are:

1. 'Watch This BEFORE You Get a Flu Shot' (Dr. Eric Z)
2. 'Vaccine Ignorance' (Dr. John Bergman)
3. 'My Opinion on Vaccinations' (Dr. Eric Berg DC)

Its top recommendation:

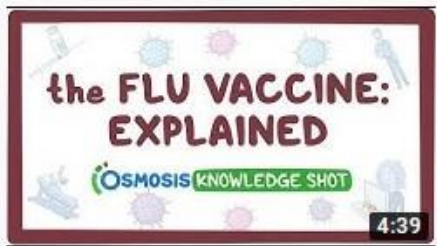
1. 'Dr. Oz No Flu Shots For My Kids!' (ExperimentalVaccines)
2. 'A Message for the Anti-Vaccine Movement' (Jimmy Kimmel Live)
3. 'Flu Shot Ingredients List' (SSFCstudybuddy)

Next recommendations:

1. 'THE UGLY TRUTH ABOUT THE FLU SHOT' (James & Lea D)
2. 'Hydrogen Peroxide – Simple Trick to Treat the Cold or Flu' (Mercola)
3. 'Flu Vaccine Exposed - Piers Morgan Struck Down - Breaking News Provides Evidence' (SubliminalProof)

WSJ (7.03.18) **YouTube** video search for “flu vaccine” generated a list of anti-vaxxer content linked to further non-scientific claims. WSJ noted the active role of **YouTube algorithms** in promoting links to increasingly extreme and contentious content.





The flu vaccine: explained
Osmosis ✓ 24K views • 11 months ago
What is the flu vaccine? The flu vaccine protects you from getting influenza viruses which can give you a high fever, runny ...
CC



Flu Vaccine: Myths and Facts | UCLA Health
UCLA Health • 35K views • 1 year ago
Learn more: <https://www.uclahealth.org/flu-prevention>.



A Shot Worth Taking: Debunking Flu Vaccine Myths (PKG)
Cleveland Clinic ✓ 15K views • 2 years ago
It's that time of the year - your doctor has probably already reminded you to get your annual flu shot. But many wonder, is it really ...



Duke Dives Deeper into Influenza Vaccine
Duke University ✓ 1.6K views • 2 weeks ago
As part of a massive national effort to improve and modernize flu shots, the Duke Human Vaccine Institute has received three ...
CC



Mayo Clinic Minute: The facts about 3 flu vaccine myths
Mayo Clinic ✓ 29K views • 10 months ago
Every influenza season, millions of Americans decide to skip the flu shot based on false information. Dr. Gregory Poland, Director ...



**Elizabeth Warren**
Sponsored • Paid for by Warren for President

Breaking news: Mark Zuckerberg and Facebook just endorsed Donald Trump for re-election.

You're probably shocked, and you might be thinking, "How could this be true?"

Well, it's not (sorry.) But it's true. Facebook didn't even let Trump free rein to lie on his platform, and then it paid Facebook's ad money to push their lies to American voters.

If Trump tries to lie in a TV ad, most networks will refuse to air it. Facebook just cashes Trump's checks.

Facebook already helped elect Donald Trump once. Now they're deliberately allowing a candidate to lie to you about the world. It's time to hold Mark Zuckerberg accountable — and I think you agree.



Mark Zuckerberg just endorsed Donald Trump
It's time to break up our biggest tech companies like Amazon, Google, and Facebook.


[Sign Up](#)

[MY.ELIZABETHWARREN.COM](#)

About the disclaimer

When an advertiser categorizes their ad as being about social issues, elections or politics, they are required to disclose who paid for the ad. [Learn more](#)

FAKE NEWS




**Elizabeth Warren**
- 14 hrs · 🌐

Facebook changed their ads policy to allow politicians to run ads with known lies—explicitly turning the platform into a disinformation-for-profit machine. This week, we decided to see just how far it goes.





We intentionally made a Facebook ad with false claims and submitted it to Facebook's ad platform to see if it'd be approved. It got approved quickly and the ad is now running on Facebook—take a look.

Facebook holds incredible power to affect elections and our national debate. They've decided to let political figures lie to you—even about Facebook itself—while their executives and their investors get even richer off the ads containing these lies. Once again, we're seeing Facebook throw its hands up to battling misinformation in the political discourse, because when profit comes up against protecting democracy, Facebook chooses profit.


The Trump campaign is currently spending \$1 million a *week* on ads including ones containing known lies—ads that TV stations refuse to air because they're false. Facebook just takes the cash, no questions asked.

 37K

3.9K comments
10K shares

 Like  Comment  Share 

Most relevant ▾

 Author

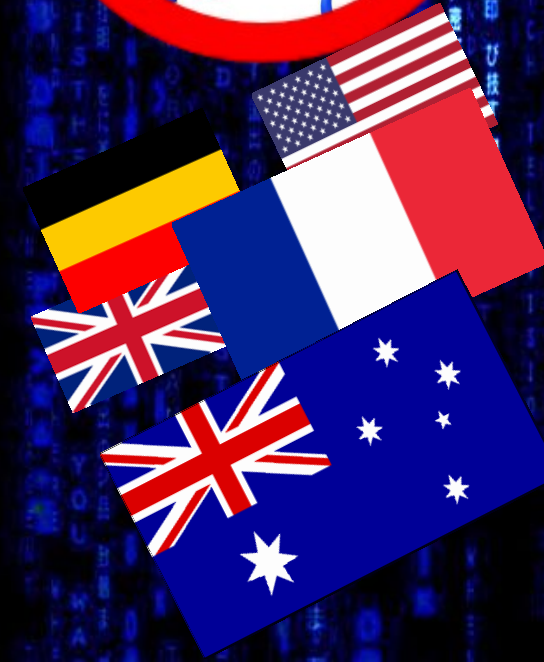
Meanwhile, just last week, US Senator Elizabeth Warren protested against Facebook's continuing complicity in the proliferation of fake news by paying for a FB advertisement proclaiming that Mark Zuckerberg had endorsed Donald Trump (with a disclaimer).



- In July 2019, Facebook was hit with a **\$5 billion fine** from the US **Federal Trade Commission** for its role in the *Cambridge Analytica* scandal including allowing abuses of personal data and phone numbers.
- In 2017 Germany introduced the **Network Enforcement Act** (NetzDG) requiring *removal of hate speech within 24 hours* and imposes fines for failing to respond (1/6 of Facebook's content moderation staff are employed in Germany).
- France and the UK have both announced **levies on the domestic turnover** of digital intermediaries (3% in those earning >€750m in France, and from 2020, 2% on companies generating >£500m in the UK). The EC is also debating a EU-wide levy while NZ is looking into a 3% Digital Services Tax.
- The UK **Online Harms White Paper** (2019) proposes a “**duty of care**” for digital intermediaries under a new regulatory body. The focus is on harmful content but extends to **source transparency** and reducing **fake news proliferation and filter-bubbles**.



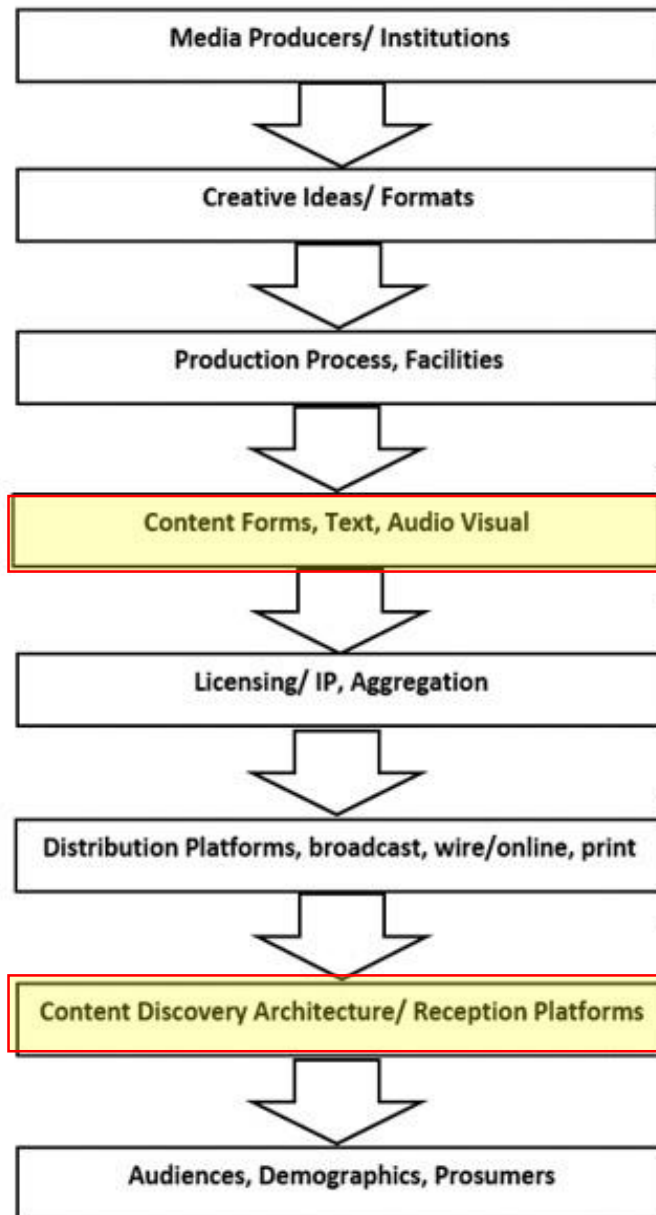
- A May 2019 French government report based on a 6 month investigation of Facebook recommended-
 - establishment of an independent regulator
 - Greater **algorithmic transparency**/ accountability of content discovery;
 - Increased responsibility for content moderation; and protection of user integrity (= **duty of care/ information fiduciary**)
- In July 2019 the French Parliament also approved a bill imposing a **24-hour take-down** limit for “obviously hateful” content.
- In Australia, the **Unlawful Showing of Abhorrent Violent Material Bill** took only 2 weeks to be approved by Parliament. Failure to expedite removal of extreme content incurs a fine up to 10% of turnover and even 3 years’ imprisonment.
- The July 2019 ACCC **Digital Platforms Inquiry** recommended rebalancing of relations between digital intermediaries and news media along with measures for greater **privacy protections** and reduction of **fake news/disinformation**.



- Facebook and the other digital intermediaries are not inviting regulation out of any sense of *civic obligation or remorse* for recent events. Their cooperative presence at the Christchurch Call is strategic and *primarily self-interested*.
 - Without legal definitions of obligations and liabilities in regard to content management, it is difficult to assess **regulatory risk**-claiming legal compliance is a convenient catch-all defence.
 - Consistent **multilateral regulatory frameworks** are less complex to manage than an accumulation of *disparate regimes/models* imposed across different jurisdictions.
 - Engagement with state regulators gives digital intermediaries a say over the shape/scope of interventions.
 - **Quarantining regulatory responses** to matters of *content moderation* does little to address the **structural network power** of the incumbent intermediaries over *content discovery* and harvesting of *personal data*.



Value Chain model of regulatory intervention points



Digital intermediaries are *neither publishers nor pipelines* providers- they occupy a space on the value chain that falls through the cracks of traditional media regulation frameworks.

The focus of the Christchurch Call is primarily on **enhancing content moderation**- this is important but it's the tip of the iceberg if the wider goal is to address **structural power of digital intermediaries**.

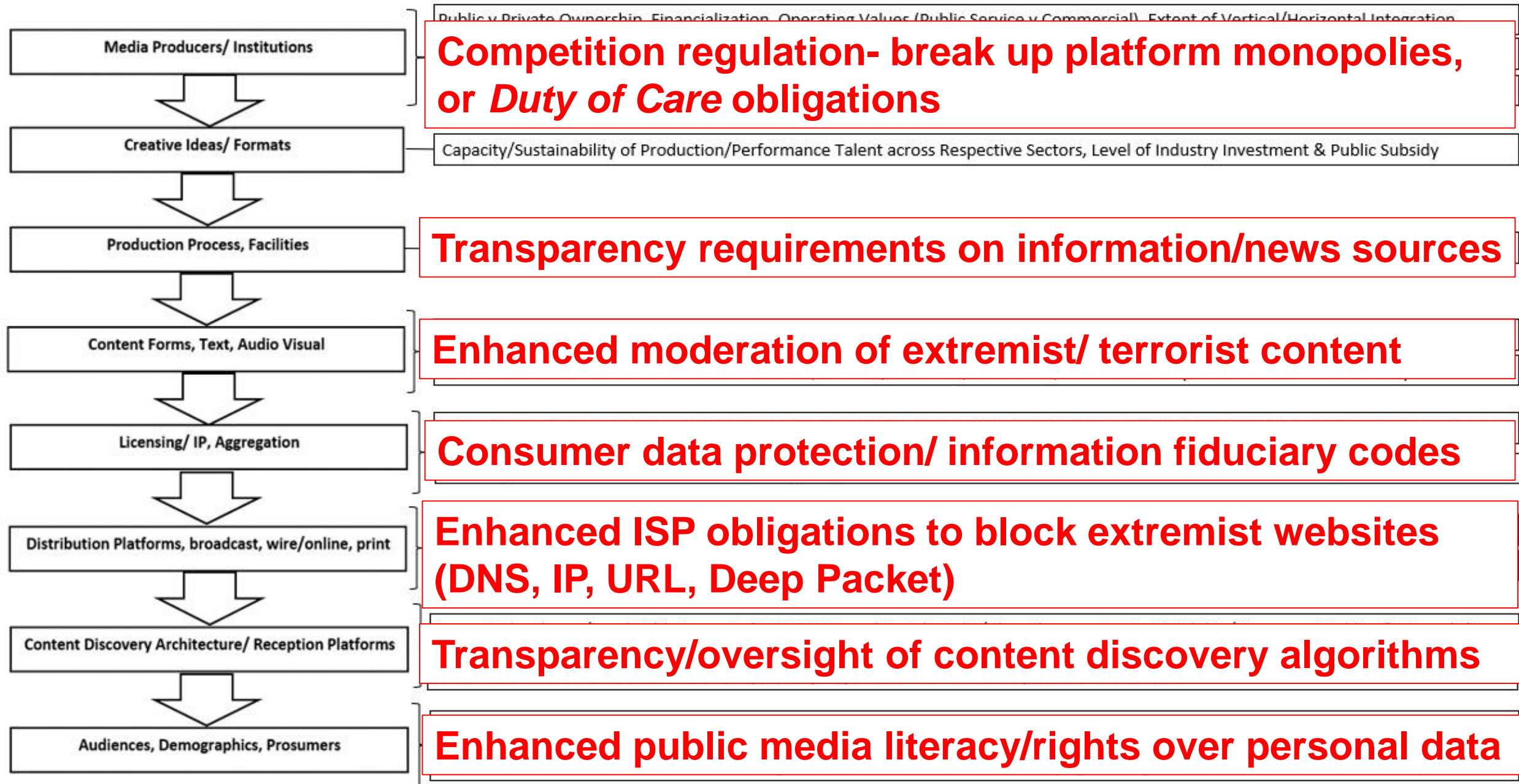
Existing regulatory frameworks can address many *content* issues- e.g. Chief censor classified terrorist video & manifesto as objectionable (one man jailed in June 2019). Several ISPs recently acted unilaterally to block **8chan**.

But neither broadcast/film & literature nor telecommunication laws adequately address the digital intermediary operations as the **platforms** providing the **algorithms/ architectures of content discovery**.

Value Chain model of regulatory intervention points



Value Chain model of regulatory intervention points



Regulatory implications for the public sector

- The regulatory frameworks for digital media intermediaries are **3 decades out of date**- broadcasting, film, telecoms have never been aligned. *Changes in government de-rail long-term planning.*
- Dealing with the policy issues raised by Christchurch, El Paso, or Halle goes beyond simply controlling extremist content- *network power of global tech firms, erosion of news media business models, proliferation of fake news/disinformation, etc.*
- The ubiquitous permeation of digital media into every facet of social activity- politics, business, culture- raises a range of **intersecting policy issues across multiple ministerial portfolio boundaries** (e.g. DIA -objectionable content/ Justice-hate-speech).
- Policy responses in one Ministry aimed at one aspect of social media may *complicate or foreclose options* available to other portfolios (e.g. IR digital services tax on intermediaries solves a *revenue* issue but that could affect tech sector investment or preclude a wider media levy to support declining news media).



Concluding points

- The Christchurch Call summit provides a useful basis to progress multilateral-supranational deliberations on regulation of digital intermediaries in the longer-term.
- Controlling extremist content proliferation on social media is important- even if the **dark web** can still enable sharing, *reducing unintentional discovery/exposure* to harmful material is vital.
- Although the major tech firms are *global*, **domestic policy responses** are still crucial- the pressure of local/regional interventions motivates intermediaries to cooperate at supra-national level. Existing domestic provisions can provide the basis for a joined-up supranational framework.
- *Avoid limiting regulation to content moderation-* this is just the tip of the iceberg. There are underlying issues of **structural network power** with *social, cultural and democratic* implications.



- Need a **wider framework of *regulatory options*** to address the *structural power* of digital intermediaries over data and content discovery-
 - New layers/points of intervention
 - New/ better-focused regulatory agents
 - New levers/mechanisms
- Need a collaborative, harmonised approach to policy formulation across disparate government departments/regulatory bodies to avoid policy fragmentation/foreclosure- spirit of **Kaitiakitanga**?
- The Christchurch Call is *not just a technical matter* which the tech sector can resolve on our behalf with a better algorithm. Need to recognise the *vested interests* of digital intermediaries which underpin current lobbying efforts and engagement:
Beware ***policy quarantining***, **NIMBlism*** and geeks bearing GIFCTs!

*Not in my back internet



Related article links:

Thompson, P. A. (2019) Beware of geeks bearing gifts: Assessing the regulatory response to the Christchurch Call . *The Political Economy of Communication* Vol 7.1 (August) Available at: <https://www.polecom.org/index.php/polecom/article/view/105/314>

Ribiero, M.H., Ottoni, R., West, R., Almeida, V.A.F. & Meira Jr, W. (2019) Auditing Radicalization Pathways on YouTube. *Federal University of Minas Gerais Research Paper*. 3 September. Available at: <https://arxiv.org/pdf/1908.08313.pdf>

Evans, R. (2018) From Memes to Infowars: How 75 Fascist Activists Were “Red-Pilled”. *Bellingcat*, 11 October. Available at: <https://www.bellingcat.com/news/americas/2018/10/11/memes-infowars-75-fascist-activists-red-pilled/>

Berger, J.M. (2018) The Alt-Right Twitter Census: Defining and describing the audience for Alt-right content on Twitter. *Vox-Pol* . Available at: https://www.voxpol.eu/download/vox-pol_publication/AltRightTwitterCensus.pdf

Presenter Email- peter.thompson@vuw.ac.nz