

A digital duty of care

Doteveryone's perspective

Jacob Ohrvik-Stott and Catherine Miller
February 2019



As the government prepares to publish its online harms white paper, the concept of a duty of care - an obligation on digital platforms and services to ensure people's safety and wellbeing - has been raised as a way of safeguarding users and regulating harmful activities online.

Digital Minister Margot James said the white paper will “*ensure that the platforms remove illegal content and prioritise the protection of users,*”¹ while Tom Watson, Shadow Secretary of State for Digital, Culture, Media and Sport also backed a duty of care saying it must “*offer meaningful protection and breaches of the legal duty must be met by robust penalties*”.²

In October, Doteveryone called for an Office for Responsible Technology³ that will give regulators across the landscape the capacity to hold technology to account, inform the public and policymakers with robust evidence and support people to seek redress from technology-driven harms.

¹ <https://www.gov.uk/government/speeches/margot-james-speech-on-safer-internet-day>

² <https://labour.org.uk/press/tom-watson-speech-fixing-distorted-digital-market/>

³ <https://doteveryone.org.uk/wp-content/uploads/2018/10/Doteveryone-Regulating-for-Responsible-Tech-Report.pdf>

Here we explore how far a duty of care will address the issues which we uncovered in our research.

What is a duty of care?

As all countries grapple with with new ways to respond to the disruptive impact of digital technologies, duty of care sits in the middle of a spectrum of approaches from the strongly interventionist - like Germany's NetzDG law that fines social media companies that don't remove untrue or criminal content within 24 hours - to the more laissez-faire approach of the United States where freedom of speech takes precedence.

So far untested internationally, if the UK pursues legislation based around a digital duty of care for online harms it may set a precedent which other countries choose to follow.

The idea of a duty of care for a digital context has been developed by William Perrin and Professor Lorna Woods for Carnegie UK Trust⁴. It has widespread support, including from the Science and Technology Select Committee⁵ and the Children's Commissioner⁶ and has been championed by the Daily Telegraph.⁷

The concept draws from existing duties of care for the offline world such as health and safety legislation and environmental regulation. These place obligations on people or organisations to take reasonable steps to ensure their activities don't cause significant harm to people, animals and the environment.

In the digital context, a duty of care aims to tackle harmful activities on digital platforms and services, particularly social media. The 5Rights Foundation have categorised harms that could be in scope as:

- **Content** - exposure to harmful material
- **Contract** - exposure to inappropriate commercial pressures
- **Contact** - comes into contact with malign actors
- **Conduct** - peer-to-peer harms, cyberbullying etc.⁸

Platforms and services would be responsible for identifying harms to users and wider society, develop measures to tackle them, and improve them if levels of harm don't fall. If they fail to meet the required standard, regulators will work with them to develop better interventions and, as a last resort, deliver fines and punishments.

⁴ <https://www.carnegieuktrust.org.uk/blog/reducing-harm-social-media-duty-care/>

⁵ <https://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/news-parliament-2017/impact-of-social-media-young-people-report-published-17-19/>

⁶ <https://www.childrenscommissioner.gov.uk/wp-content/uploads/2019/02/cco-duty-of-care-owed-by-online-service-providers-to-children.pdf>

⁷ <https://www.telegraph.co.uk/news/0/duty-care-campaign/>

⁸ <https://d1qmdf3vop2l07.cloudfront.net/eggplant-cherry.cloudvent.net/compressed/c1c90b5e630cd6bac92b6ad41afab22d.pdf>

How far does a duty of care address the need for regulation?

The slew of revelations about the negative impacts of technology have quickly moved the debate in the UK from whether to regulate digital platforms and services to how to regulate them. Most recently Molly Russell's suicide⁹, which her father in part attributes to material she encountered on Instagram - has created an emotive and urgent political context. This generates useful momentum for government to act. But there's also a risk that it distracts from larger issues and that policymakers will indulge in populist gestures for short term gain with little long term impact.

It's important to remember that a duty of care is only designed to address one small part of the current gaps in the regulatory landscape. As Doteveryone highlighted in *Regulating for Responsible Technology*, all regulators across all sectors are struggling to have the remit, capacity and evidence base to address the disruptive impacts of digital technologies.

Without a coherent response to the underlying business models of technology, the algorithmic processes and design patterns within technology and the impacts of technology on social infrastructure, a duty of care can only be a symptomatic treatment of the consequences of tech on one aspect of life.

And there's a danger that duty of care sucks up all the available political capacity for regulation and leaves the broader landscape untouched.

Doteveryone would encourage policymakers to think beyond the noisy headlines and ensure they address the fundamental changes needed to regulate in a digital age.

However, recognising the narrow scope it seeks to address, a duty of care has merits as a pragmatic and collaborative approach. It rebalances the relationship between industry and the state by giving parliament and a regulator responsibility for setting the terms for what the UK, as a society, considers harmful and wants to eradicate. And it puts the onus on business to find the mechanisms to achieve these outcomes, with penalties if it fails to do so.

Founded on the precautionary principle, the inherent flexibility of a duty of care - focusing on safety and wellbeing rather than specific uses of technology - also overcomes the problem of policymakers' tendency to legislate for yesterday's technology while engineers create new facts on the ground.

In practice, it's likely that the tensions between safeguarding and freedom of speech will be arbitrated by the platforms and services themselves. Although this is problematic to an extent, it does not greatly differ from other areas where there is already content regulation, such as in broadcasting. With sufficient requirements for transparency around these decisions, avenues for appeal, oversight and redress these questions are likely to be negotiated over time and new norms will be established.

The great challenge for any regulation is around enforcement. There's scepticism that cash-rich platforms will flinch in the face of fines and penalties. However both Germany's NetzDG and the

⁹ <https://www.bbc.co.uk/news/av/uk-46966009/instagram-helped-kill-my-daughter>

EU-wide GDPR data protection legislation have provoked platforms into action with Facebook recruiting additional moderators in Germany and Google appealing a relatively minor fine from the French data protection authorities. It's likely that for reputational reasons alone platforms would at the very least pay lip-service to a duty of care.

Putting a duty of care into practice

Duty of care is only as effective as its implementation. As policymakers consider what a duty of care should involve in practice, these are the questions Doteveryone considers important and our recommendations for dealing with them,

How can regulators look beyond outcomes?

Regulatory duties of care are expressed in terms of the outcomes they want to achieve - a reduction in levels of specified harms - without prescribing specific actions companies must take to achieve them. This gives digital services and platforms freedom to innovate, and accounts for dynamic and complex environments where regulations can't be developed for every eventuality.

But there are risks. Reporting metrics may be gamed, and companies may be penalised unfairly if a harm increases due to forces beyond their control - for example, rising hate speech could be symptom of a more fractured political climate.¹⁰

An outcomes focus is also challenging in a fast moving environment where there is little rigorous evidence base. The Royal College of Paediatrics and Child Health recently noted, for example, that the view that screen time is directly toxic to health is "*popular outside the scientific literature, but has essentially no evidence to support it*".¹¹

In this context, a duty of care should be reframed in terms of the responsibility of the processes in place and the care taken to foster positive impacts for society.

Our recommendation - scrutinise processes

Harm prevention should not be a mopping up exercise once a product has been released onto the public. Instead, it should be built into technologies upstream.

Doteveryone advocates Responsible Technology that considers the context, consequences and contributions of a technology from the outset. A duty of care should incorporate emerging best practices around responsible design into a code of practice. And the mechanisms used to address harms also need to be scrutinised - whether that's AI to tackle hate speech or a complaints handling procedure. Other measures - such as formal standards or open APIs that allow civil society to develop their own harm mitigation tools for digital services - should also be developed.

¹⁰

<https://www.theguardian.com/society/2018/oct/16/hate-crime-brexit-terrorist-attacks-england-wales>

¹¹ https://www.rcpch.ac.uk/sites/default/files/2018-12/rcpch_screen_time_guide_-_final.pdf

How will this system win public trust?

Doteveryone's *People Power and Technology*¹² surveys found the public felt disempowered and distrustful in relation to technology companies. For a duty of care to be effective, the public must believe that the regulator enforcing it is on their side - understanding and channelling their concerns.

Our recommendation: don't just regulate, reassure

Technocratic solutions are not enough. A new duty of care will need a credible and relatable leader to engage with the public and articulate how their concerns are being met. In the UK, the Information Commissioner's Office has benefited greatly from the presence of Elizabeth Denham as a visible and authoritative figurehead. In the wake of the GDPR, awareness of data and privacy rights are at a record high.¹³

How can the sector learn while ensuring genuine accountability?

Digital platforms will need to work through new and unfamiliar responsibilities to complex problems. Openness about the challenges they encounter, and the innovations they develop to mitigate digital harms, will be crucial to raise standards. Recent behaviour by platforms - Facebook in particular - has seen a tendency to deny and obfuscate.

Our recommendation: incentivise collaboration

Digital services must be properly held to account for the harms of their work - but they need carrots as well as sticks. For example, Companies might be shown more leniency if they flag spikes in harms to regulators at an earlier stage.

Or there could be incentives to innovate harm mitigation tools and strategies that all companies can benefit from.

Where does the balance of moderating power lie?

If implemented in a blunt way, duty of care may incentivise platforms to over-moderate users' communications and content - potentially intruding further into people's privacy, jeopardising

¹² <https://doteveryone.org.uk/project/peoplepowertech/>

¹³ <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2018/07/annual-report-2017-18/>

freedom of speech. There's a need for checks and balances that allow users to appeal decisions and others to scrutinise the moderation decisions made.

Our recommendation: establish systems of redress

To counterbalance potential overzealous moderation, the public and trusted advocacy groups must be able to contest decisions and seek redress from harms.

This could be delivered by a statutory ombudsman associated with the duty of care regulator to ensure independence and neutrality, and collective redress should be possible where a specific harm has affected a large number of people.

How should action on harms be prioritised?

The UK Government's Internet Safety Strategy Green Paper outlines a wide range of potential issues for new digital regulation to cover. They span harms to the economy, justice and democracy, national security, safeguarding of young people and users' emotional and physical well-being¹⁴.

That's a lot for one regulator to take on.

With digital technologies in a state of constant change, it's essential to consider how regulators can remain responsive to emerging harms without drowning in an ever-expanding remit and workload.

Our recommendation: look beyond user harm.

Many digital services' impacts are felt beyond their immediate users - they therefore also owe a duty of care to wider society. As well as monitoring the prevalence of harms from services, regulators should lead and collate research about the impacts of harms on society collectively - establishing an evidence-base that allows regulators to direct their attention and resources where there is greatest need.

¹⁴https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650949/Internet_Safety_Strategy_green_paper.pdf