DUTY OF CARE

There are tried and tested models from other sectors that could be adapted for social media regulation. **LORNA WOODS** puts forward a proposal

It has become increasingly apparent that, while the internet offers many possibilities for people to connect and to find and share information and knowledge, there are also problems which come with the way this technology is used – ranging from data breaches, illegal activity (child pornography, terrorism), to less serious issues, such as cyber bullying and the dissemination of fake news. There are also concerns about whether the design of services cause problems, as well as their allegedly addictive nature.

To date, the policy narrative has tended towards emphasising the benefits of the internet and the difficulties of regulating in fast moving fields. The cumulative impact suggests that market forces and a self-regulatory approach are no longer producing an ideal outcome, a point seemingly now accepted even by some of the tech companies. Mark Zuckerberg, in his evidence to the US Congress, has said he welcomes regulation of the right sort and Jack Dorsey of Twitter has made a public plea for help. The emerging evidence of harms suggests that, following the precautionary principle, there is a case for regulatory measures.

This article outlines a proposal, elaborated with Will Perrin under the aegis of the Carnegie UK Trust, to try to reduce the harms in this sector; it focuses on social media platforms, businesses which for many people have almost become the entirety of their internet use. While some of the regulatory examples are British, it is suggested that the model is adaptable to other legal systems too.

THE PUBLISHER PROBLEM

Currently, the discussion about how to treat social media platforms has been phrased in the context of the question of whether they are publishers of their users' content and whether it is therefore appropriate to hold them liable for that content. The alternative sees the platforms as intermediaries, little more than conduits for others' content. This indeed seems to be the model for many laws worldwide dealing with intermediaries, one example being the European Union's e-commerce directive,[1] which provides "safe harbour" from liability for neutral intermediaries which have no knowledge of problematic content. Once such intermediaries cease to be neutral, or have knowledge of a particular item of content, they lose immunity from suit and effectively are treated as publishers.

This all-or-nothing strategy, which predates the introduction of social media, does not fit today's world as it does not recognise the scale of the operations involved and the way that content must be organised to make it accessible, rather than

overwhelming. The current approach offers either too much protection, so that the platforms become reckless, or too little – especially where platforms are deemed not to be neutral, with the economic threats resulting from extensive liability possibly resulting in an overly cautious approach to providing space for a wide range of users and content.

We thought that, rather than consider the question as being about either transmission or content, a better framing would be to view social media as "quasi public spaces". We are not referring here to some form of digital agora or "coffee shop" where matters of public interest are discussed, but instead to the range of spaces that are generally accessible by the general public or at least a large number of people. We are thinking of shopping malls, cafes, night clubs, sports stadia, parks, children's playgrounds, and the streets and squares of our towns and cities, to give but a few examples. This change in focus allows us to recognise the fact that social media constitutes an environment provided and controlled by the operators within which users carry out a range of activities, and is significant for a number of reasons:

- Different spaces may be used for different purposes and have different user groups with specific social norms about appropriate behaviour – and consequently lead to different risks of harm arising
- The architecture of the place is significant in terms of what behaviours it facilitates, permits, prohibits and encourages – a recognition also found in Lawrence Lessig's work where "code is law"[2]
- A proposal which focuses on the underlying system – the architecture or the arena provided – allows us to decouple "safety" concerns from individual items of content.

The question then becomes how do we make these spaces safer.

MODELS FOR REGULATING THE ARENA

We considered the models that exist for broadcast content and telecoms. In the UK, these are found in the Communications Act 2003 but while they are comparatively light touch compared with the equivalent regimes in, say, the 1980s, we consider they are too directive in terms of the obligations imposed on service operators. Moreover, ultimately both regimes allow the relevant regulator to stop a company from providing its service, which we initially felt was too extreme for speech-related services. Instead, we looked at two groups of regulation: that concerning physical space/environment, and data protection (which is particularly relevant in the technology sector but has a broader application).

In the UK, the Occupiers Liability Act 1957, the Health and Safety Act 1974 and the Environmental Protection Act 1990 have in common the fact that they each establish a statutory duty of care. For example, section 2(2) of the Occupiers Liability Act specifies that "[t]he common duty of care is a duty to take such care as in all the circumstances of the case is reasonable to see that the visitor will be reasonably safe in using the premises for the purposes for which he is invited or permitted by the occupier to be there". This puts an obligation on the person controlling the space to ensure – within the limits of reasonableness – the safety of visitors.

Similar phraseology can be seen in the Health and Safety Act 1974, which applies to almost all employers and their myriad activities. The regime does not set down specific detailed rules with regards to what must be done in each workplace but rather sets out some general duties that employers have both as regards their employees and the general public. It then elaborates on particular routes by which that duty of care might be achieved: e.g. provision of machinery that is safe; the training of relevant individuals; and the maintenance of a safe working environment.

> **"We looked at two groups of regulation: physical space/environment, and data protection."**

While the Health and Safety at Work Act sets goals, it leaves employers free to determine what measures to take based on risk assessment; this allows a certain amount of flexibility. In that employers may change how they respond to risks, an element of future-proofing is also built in – measures change in line with a changing environment. This then is a counter to the risk that through shoddy practices an employer might attempt to externalise business costs (e.g. onto a worker; onto society).

In contrast to the Occupiers Liability Act, the Health and Safety at Work Act also provides for a regulator: the Health and Safety Executive (HSE), whose functions are set down in the Act. It may carry out investigations into incidents, and has the power to approve codes of conduct. It also has enforcement responsibilities and may serve "improvement notices" as well as "prohibition notices". As a last measure, the HSE may prosecute. There are sentencing guidelines which identify factors that influence the heaviness of the penalty. Matters that tend towards high penalties include:

- Flagrant disregard of the law
- Failing to adopt measures that are recognised standards
- Failing to respond to concerns
- Failure to change/review systems following a prior incident
- Serious or systematic failure within the organisation to address risk.

A similar duty of care approach can be seen in the Environmental Protection Act for disposal of waste, again with the effect that business costs (relating to the safe and appropriate disposal of waste) are not externalised. Economic theory suggests that markets do not work for society unless a business accounts for all its costs in making decisions – the ones it incurs directly and those that fall on others. Regulation often plays a role in internalising external costs through a variety of routes such as the "polluter pays" principle, which can be seen also in the other examples. In the waste disposal context,

while the legislation specifies particular concerns, more detail on what the duty of care requires is set down in secondary legislation, and codes of practice give practical guidance. The documentation demonstrating compliance with the duty of care must be kept for 2 years.

Given the importance of data protection for the tech sector, the model adopted in the EU's General Data Protection Regulation (GDPR) is also relevant.[3] Central to the GDPR is the principle of accountability, which can require a data controller, as defined in the GDPR, to show how it complied with the rules on data protection. Another approach is the risk-based approach to data protection measures, so the more sensitive the data, the great the volume of data, or the riskier the nature of the processing, the more care should be taken. The GDPR, for example, requires "appropriate technical and organisational measures to ensure a level of security appropriate to the risk".

Specifically, the GDPR requires "data protection by design and by default" (privacy by design),[4] so that the protection of personal data is a default property of systems and services. The design includes business practices as well as technology choices. Additionally, the GDPR envisages that data controllers should carry out privacy impact assessments (PIAs), at least for certain forms of "risky" processing[5] – though it is generally seen as good practice for all large processing operations. Guidance on how to carry out these requirements will be given by the relevant independent regulatory authority. Article 34 specifies that where a PIA suggests that there is a high risk, the controller must ask the supervisory authority before proceeding. The purpose of such an assessment is to help a controller first identify risks and then mitigate against them.

So while accepting the risk-based approach, the purpose is also to stop the problem happening in the first place, and might be said to constitute a precautionary approach. A similarly precautionary approach to harm can be seen in the requirements of privacy by design and security by design. These themes feed into the factors that the supervisory authorities take into account when assessing the size of fines to impose on a controller (or processor) in breach under the GDPR.

SALIENT POINTS

A number of points can be drawn from these regimes. The first concerns the advantages of a duty of care that targets the systems that the platform provider operates, rather than focusing on individual instances of harm. The reliance on a systems-based approach may mean that instances of harm are less likely to happen in the first place; systems based on cases of harm depend on those instances having occurred.

Although general obligations are identified, the responsibility lies with the operator to understand the risks involved in its operations and to take appropriate action to protect against them. This allows flexibility in the measures that are implemented, and operators may to some extent choose approaches that they feel suit them best, mitigating to some extent the concern that safety requirements operate as a barrier to entry.

It also means that the measures could keep up to date with developments in technologies and in the market more easily than legislatively mandated solutions. To a certain extent, a duty of care thus allows for future-proofing of the safeguarding system.

High risk activities in these regimes are, however, subject to tighter control and even in some instances a "permissioning" regime, so that the activity cannot go ahead without consent of the regulator. The data protection and health and safety regimes highlight that there may be differing risks with two consequences:

- That measures taken to prevent risks actually happening should be proportionate to those risks
- That in areas of greater risk there may be greater oversight.

Nonetheless, while the data protection regime may impose – post GDPR – hefty penalties, and limit particular processing operations, it may not stop a controller from being a controller. Again, with regard to health and safety, particular activities may be the subject of a prohibition notice, but this does not disqualify the recipient from being an employer. The notice instead relates to a particular behaviour.

Most of the regimes have some form of oversight and enforcement – including criminal penalties in cases of flagrant breach. The health and safety regime gives some examples (noted above) of the types of factors that could be taken into account when considering penalties. Criminal action in extreme cases can be seen in the other sectors.

> **A duty of care can target the systems rather than focusing on individual instances of harm.**

A recent example comes from data protection in the UK where the Information Commissioner is starting criminal prosecution against a data controller for failure to comply with an enforcement notice to comply with a subject access request.[6]

The regulators responsible are independent from both parliament and industry and have the authority to make decisions without political approval. In some sectors, such as broadcasting, there are standards for independence.[7] Any regulatory scheme should consider these elements. One key element of many regulatory approaches is that changes in policy take place in a transparent manner after consultation with a range of stakeholders and on the basis of (empirical) evidence.

We have developed a checklist of elements (see overleaf) that a regulatory schema should satisfy and which we believe our proposal below, which is built on the models discussed, incorporates.

THE PROPOSAL

Drawing on these models, we propose imposing a duty of care on social media platforms to take reasonable steps to protect against foreseeable risk of harms to their respective users. A duty of care

CHECKLIST FOR REGULATION

Criteria	What we mean
1 Proportionate	• Regulation will not apply to every platform, with prioritisation of largest providers • Adaptable to specific circumstances, behaviours, cultures and users • Resource input can be justified by the outcomes
2 Future-proof, flexible, iterative	• Outcomes/principles based – not a set of detailed/prescriptive rules • Can be quickly applied to new platforms and new types of harm • Principle of safety by design through system level regulation, not regulation of specific content • Co-regulatory approach, with providers identifying and implementing solutions to tackle harms • Drives process of continuous improvement
3 Risk-based	• Appropriate consideration of the intended and unintended consequences • Allows for differentiation between different types of audience, including extra protection for more vulnerable groups
4 Independent	• Neutral of political or commercial interest
5 Consistent	• Clear process and criteria for qualifying social media services • Actions can be measured
6 Transparent	• Process is open and understood, particularly by the public • Actions are published, known and in the public domain • Consultative process with many groups, including the hardest to reach
7 Tested	• Regulatory model draws on existing concepts and proven models (which demonstrate longevity) on scope, quantity and impact • Regulator has sufficient industry skills and expertise to act, such as ability to employ at competitive rates people who have worked in relevant industry roles
8 Simple	• Utilises existing resources where possible, such as use of existing regulator • Relatively straightforward to legislate for
9 Promotes market efficiency	• Ensures that external costs are brought into a company's production decision along the lines of the "polluter pays" principle • Incentivises/rewards good practice • Does not act as a barrier to entry
10 Enforceable	• Backed by statute • Sanctions are feasible in terms of implementation, proportionate, provide a genuine deterrent and offer a process for escalation
11 Interoperable, scalable	• Offers potential to operate effectively in an international context

requires the operator to act towards its users with the attention, caution and prudence that a reasonable person in the circumstances would. It is not a requirement to protect users from all possible harms but instead for the operator to consider its own operations and their likely consequences (taking into account the likely behaviour of its users) and to take reasonable steps to guard against those matters that can be reasonably foreseen.

In terms of foreseeability, not every possible harm is a foreseeable harm; conversely, neither is wilful blindness on the part of the operator acceptable. In contrast to the English common law duty of care, which is limited insofar as mental and emotional harms are concerned, we propose that the relevant harms here should in principle include such emotional harms.

While the obligation to prevent harm is set down in statute, we envisage setting up an independent regulatory body, and one of its tasks would be to lay down further guidance as to what those harms look like. In so doing, we envisage that it would consider existing evidence and consult relevant stakeholders (e.g. industry, NGOs, consumer bodies, service users). It could also constitute a venue for sharing best practice, development of standards and assessing rising threats. It has been noted in the context of cybersecurity that such a venue is important because no one industry player can see the entire information ecosystem.

Affected companies would assess their platforms, technologies and services for risk of harms, and measure them and draw up a plan to mitigate those risks. These proposals could include, depending on the service:

- Enforcement of terms of use/community standards equally and transparently
- Parental controls
- Tools to find or filter content that are easy to use and in the control of users
- An effective complaints mechanism (both as regards other users but also the company)
- Contributions to a mediation service
- A review of mechanisms for prioritising content.

It would be expected that for new services, amendments to existing services and new technologies, a "harms impact assessment" should be undertaken, and that companies should adopt a safety by design approach – safety understood in the context of their user group (so greater if the service is aimed at children, for example). It may be that for smaller companies, the regulator can develop more detailed guidance/advice in this area.

The regulatory body would then keep progress under review to check whether the companies are doing what they said and whether what they propose is effective (see harm reduction cycle diagram) with the company reviewing its plans in the light of experience. Note the focus is on the systems – whether software, or business (including legal issues) – that the operator has in place, so instances of "bad content" would not automatically mean that there had been a failure in the duty of care, though frequent failings might indicate just such a systematic failing.

To facilitate this, companies would be required to report (in a form set by the regulator), and the regulator would have power to compel companies to provide information, and even to raid premises in cases of urgency, subject to judicial approval. We also suggest that the regulatory body has enforcement powers – so able to issue "improvement notices" and "prohibition notices", as well as levy fines on the scale of those found in the GDPR. The regulator should also be obliged to report to government on its activities.

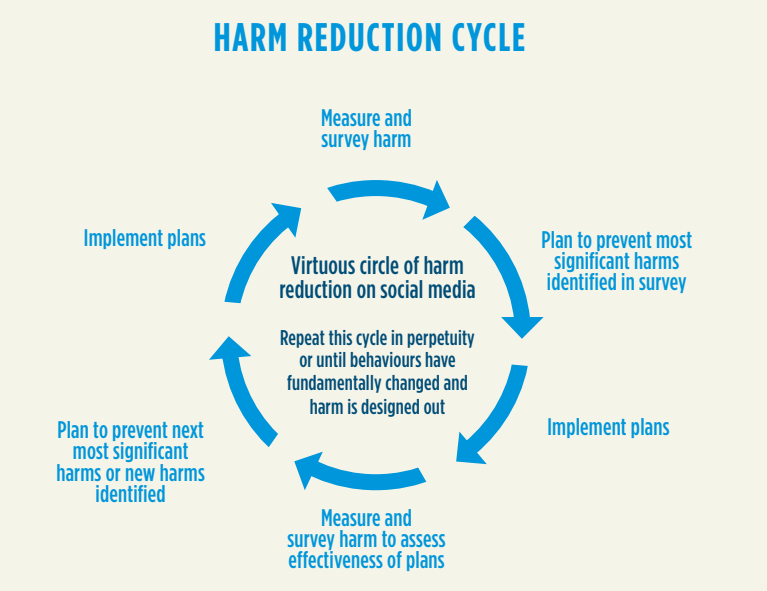
Essentially, what is proposed is a form of co-regulation, with companies setting their own standards, possibly developing their own codes of practice, within a statutory framework under the oversight of a regulator. We consider, however, that it would not be appropriate to allow a user to sue the relevant companies for breach of statutory duty (matching the position now with regard to the UK Health and Safety Act); this would not affect any other claim an individual might bring should the facts in an individual case support such a cause of action (e.g. negligence, defamation, misuse of private information).

One final question is which companies should be caught. Our proposal focuses on social media, not other actors, for example search engines, involved in the internet. We leave open the question of whether such a duty of care should be applied to these others. Who then are social media providers? We propose the following, that to fall within the regime a service should:

- Have a strong two-way or multiway communications component
- Display and organise user generated content publicly or to a large member/user audience
- Not already be subject to a sector specific regime (e.g. mass media).

This test would be assessed by reference to users in the relevant jurisdiction. We also wondered about instituting a de minimis threshold for the regulation, but have been persuaded that this is not likely to be tenable. Apps can increase in popularity so fast that designers should be thinking about baking safety by design into their software and business practices from the beginning; moreover, there are some groups which are sufficiently vulnerable (e.g. children) that any business aiming a service at them should take an appropriate level of care no matter what its size. It seems that a proportionality test is a better mechanism by which to take into account the position of small companies, so that their size and limited resources are factors taken into account when assessing the suitability of the harm prevention systems such companies propose.

The more difficult questions relate to what to do in extreme cases. Should there be a power to send a social media services company director to prison or to turn off the service? Regulation of health and safety in the UK allows the regulator in extreme circumstances, which often involve a death or repeated, persistent breaches, to seek a custodial sentence for a director. The UK's Digital Economy Act contains powers[8] for the age verification regulator to issue a notice to internet service providers to block a website in the UK. Should there be equivalent powers to send a social media services company director to prison or to turn off the service? In the US, the new FOSTA-SESTA package[9] apparently provides for criminal penalties (including we think arrest) for internet companies that facilitate sex trafficking. Given the impact on freedom of expression, such penalties should be imposed only in the most extreme cases; but should they be there at all?



CONCLUSION

The public debate is often about individual items of content and the specific rules that it might or might not have broken. This debate sometimes misses the need for an overarching duty of care to apply to the systems in place. The advantage of this approach is that to a large extent it decouples the question of whether the system operator has done a "good job" from the questions surrounding how their users use the platform. This is not to say that there will never be discussions about the type of content made available via social media, or that this proposal is a silver bullet dealing with all forms of harm. Rather, we see a statutory duty of care enforced by a regulator as providing a foundation for more detailed interventions that governments might think necessary to protect the public; indeed, a duty of care might well reveal where more sophisticated rules are necessary.

Statutory duties of care enforced by a regulator have proven to be robust and effective in the UK in a wide range of circumstances, including the complex and economically and socially profound. We see no reason why the issues posed by social media should lie beyond the capabilities of this sort of model and therefore submit that this model is worthy of consideration in the social media space.

LORNA WOODS *is professor of law at the University of Essex, UK. The proposal is under development under the aegis of the Carnegie UK Trust, and suggestions on improving it are welcome at comms@carnegieuk.org; see also blogs on the project at bit.ly/2PK1qAB.*

REFERENCES **1** Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the internal market (E-Commerce Directive) [2000]. OJ L 178/1. bit.ly/2AsiwKM **2** Lessig L (1999). Code and Other Laws of Cyberspace. Basic Books. **3** Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation (GDPR)) [2016]. OJ L 119/1. bit.ly/2vHVeNC **4** Article 25, GDPR **5** Article 33, GDPR **6** Information Commissioner's Office (2018). Investigation into the use of data analytics in political campaigns: Investigation update. 11 July. bit.ly/2KOJ67V **7** See e.g. Council of Europe Recommendation Rec (2000) 23 of the Committee of Ministers to member states on the independence and functions of regulatory authorities for the broadcasting sector. bit.ly/2Pjhycj and contrast with: International Nuclear Safety Advisory Group (2003). Independence in regulatory decision-making (INSAG-17). bit.ly/2CFY7n7 **8** Section 23, Digital Economy Act 2017. bit.ly/2qbYYo7 **9** Romani A (2018). A new law intended to curb sex trafficking threatens the future of the internet as we know it. Vox, 18 April. bit.ly/2EK1qqE