What principles should guide policy-makers in designing local and international approaches for internet intermediaries in the evolving communications environment?

Executive Summary

The Internet means different things to different people. To me personally, the Internet and the intermediary services that have been built atop it represent opportunity: the opportunity to build new connections across cultures and to express my identity. Countless others have experienced similar benefits. But for some, the risks of intermediary services have come to outweigh the opportunities. Given these differences in individuals' experience, what objectives should policy-makers aim towards as they design local and international approaches for intermediary services?

In this report, I argue that existing intermediary regimes' objectives of protecting against harm, promoting free expression, and encouraging innovation remain relevant, but that an overarching objective should be to strengthen trust. It is neither practical nor desirable to aim for an Internet in which we eradicate the potential for harm, and to do so would risk significantly impacting on the benefits that intermediaries provide. This is especially the case in view of the continuous evolution of intermediaries and the policy issues relating to them. A more realistic measure of success is that individuals and societies trust that, firstly, there is a fair and accountable framework for balancing the objectives of protecting against harm and promoting free expression and, secondly, the framework can respond effectively when new issues arise.

I outline the following six principles to guide policy-makers in their pursuit of this objective:

1) **Clarity** – Policy-makers should provide clarity on the objectives of their approaches to internet intermediaries and on the roles and responsibilities of different stakeholders.
2) **Adaptability** – Approaches should be designed so that they can adapt to the fast-changing nature of internet intermediaries themselves and the policy issues surrounding them.
3) **Accountability** – Approaches should strengthen accountability for governments, internet intermediaries and users, as well as accountability for how the system is working as a whole.
4) **Empowerment** – Approaches should aim to increase individual users' choice and control in relation to internet intermediaries.
5) **Practicality** – Approaches should take account of the commercial impact and should preserve flexibility for continuous innovation by internet intermediaries.
6) **Cooperation** – Approaches should provide for transparent, inclusive, and effective cooperation between relevant stakeholders.

Though principles are important, they can only take us so far. Policy-makers are confronting the complex trade-offs that arise in regulating intermediaries, particularly how far to go in protecting against harm without disproportionately limiting free speech and intermediaries' ability to innovate to provide benefit for their users. In this report, I aim to make these principles practical and useful, by explaining how they can help address some of these trade-offs and contribute to strengthening trust in the online

environment. I also outline some of the potential tensions between these principles and address how those tensions can be mitigated.

Background

The existing framework of intermediary liability laws was, broadly, established in the pursuit of three aims: firstly, protecting online users against harm; secondly, promoting freedom of expression; thirdly, enabling innovation and economic growth.

However, governments and publics in many countries increasingly believe the current frameworks require updating. Their most commonly stated concerns include: firstly, the proliferation of illegal content and activity online; secondly, the spread and potential for virality of content that is legal but may be harmful, such as disinformation around COVID-19 vaccines; thirdly, that intermediaries are insufficiently accountable for the power they exercise over users' freedom of expression.

Different organisations are pursuing a number of approaches to address these concerns:

- **Government regulation** – The EU, Germany, Ireland, France, US, UK, Canada, Australia, India have all introduced or are actively considering legislation introducing new responsibilities for internet intermediaries. All aim to distinguish their approach from that of China, where many foreign-owned intermediaries are blocked and Chinese-owned services are required to censor politically contentious content.[1]
- **Industry initiatives** – Internet intermediaries have invested in automated and human content moderation, published transparency reports on content moderation and, in the case of Facebook, established an independent Oversight Board responsible for scrutinising content moderation decisions.[2]
- **Civil society and inter-governmental initiatives** – Civil society groups and academics have contributed policy statements around how Internet governance can protect freedom of expression, including the Manila Principles and the Santa Clara Principles.[3]

There have been incremental steps towards international cooperation on defining a common approach to internet intermediaries. In April 2021, G7 countries, along with South Korea, Australia and South Africa, agreed on a set of Internet Safety Principles, which include broad language around the need for multi-stakeholder cooperation, fostering human rights online, and increasing intermediaries' accountability.[4]

However, there remains a lack of consensus over how approaches to intermediaries should function in practice. Even within the EU, in 2019 the European Commission expressed opposition to a French proposal to require platforms to remove notified

[1] https://daxueconsulting.com/internet-censorship-in-china/.
[2] https://oversightboard.com/.
[3] https://manilaprinciples.org/index.html; https://santaclaraprinciples.org/.
[4] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/986161/Annex_3__Internet_Safety_Principles.pdf.

content within 24 hours, outlining concerns that this would lead to "an excessive removal of content".[5]

Objectives

As a first step, policy-makers should develop realistic and concrete objectives. Politicians often imply that their proposal for regulating online intermediaries would have stopped a particular issue from arising. But in doing so, they risk setting expectations that are impossible to meet. For example, in January 2021, Canadian heritage ministry spokesperson, Camille Gagné-Raynauld, referred to the US Capitol Hill riots earlier that month as a "concrete example" for assessing the "potential effectiveness" of Canada's planned online safety regulation.[6] But most of the relevant online content linked to the riots – for instance, content challenging the legitimacy of the US Presidential Election result - was legal, and it is unclear how Canada's planned regulation, aimed primarily at illegal content, would address this material.

Policy-makers should place greater emphasis on the objective of **strengthening trust**. A central objective should be to strengthen the trust of individual users, individual societies, and countries cooperating internationally that there is a fair and accountable framework for balancing competing objectives and, secondly, the framework can respond effectively when new issues arise.

In Edelman's global 2021 Trust Barometer, trust in the social media sector was the lowest of any sector.[7] Internet intermediaries now have an important influence on individuals and our societies. This means that trust in those services is important, firstly, for individual well-being: as one example, Amnesty International research has found that women who doubted intermediaries' ability to protect them from online abuse were more likely to self-censor and withdraw from online spaces, depriving them of the potential benefits those spaces provide.[8]

Trust in intermediaries is also important for wider social cohesion. For instance, the perception of certain groups that content moderation is weighted against them has fuelled their sense of disaffection and marginalisation. This includes groups as diverse as US Republican voters and LGBT+ activists in France.[9] Intermediary approaches cannot themselves solve societal issues such as the polarisation of political perspectives, but they can aim to restore the trust that different individuals and groups place in the fairness and integrity of the online environment.

Beneath this overarching objective of trust, the three original objectives of intermediary liability laws – **protecting against harm**, **safeguarding free expression**, **promoting innovation** – should remain a central focus.

---

[5] https://www.contexte.com/article/numerique/document-les-severes-observations-de-la-commission-sur-la-ppl-avia_108449.html.
[6] https://www.thestar.com/politics/federal/2021/01/21/after-the-capitol-riots-ottawa-draws-lessons-about-social-media-regulation.html.
[7] https://www.edelman.com/trust/2021-trust-barometer, p46 of report.
[8] https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-5/.
[9] https://www.pewresearch.org/internet/2020/08/19/most-americans-think-social-media-sites-censor-political-viewpoints/; https://tetu.com/2020/05/13/les-associations-lgbt-inquietes-apres-le-vote-de-la-loi-avia-contre-la-haine-en-ligne/.

However, each individual and each society brings a different understanding of what the right balance is in protecting against harm and promoting free expression. Automated content moderation tools, for example, may be effective in identifying some forms of illegal content at scale, but do not have a 100% accuracy rate and may lead to the inadvertent removal of legal content without appropriate human oversight.[10] The European Commission's draft Digital Services Act does not include a requirement to use automated tools, whereas India's draft Information Technology Rules does require intermediaries to deploy these tools.[11] The subjectivity of concepts such as harm and the complexity of debates around online content means that trust that there is a fair and accountable framework for making these trade-offs should itself be a central objective.

This objective of trust needs to be dynamic: it needs to be maintained even as the nature of the online communications environment evolves. A prescient example is Facebook's shift from "town squares" to "living rooms", the terms used by Mark Zuckerberg to denote an increased emphasis on private, end-to-end encrypted spaces.[12] This presents an important challenge to policy-makers, as they consider what an appropriate balance of protecting against harm and protecting users' rights – including privacy, in particular – should be in more private spaces.

Principles

1) **Clarity** – Policy-makers should provide clarity on the objectives of their approaches to internet intermediaries and on the roles and responsibilities of different stakeholders.

Policy-makers need to clearly outline specific and well-defined objectives and the relationship between them. Obligations for intermediaries should be as clear as possible to reduce ambiguity, balancing this with maintaining a framework that can adapt to the evolution of the online environment. Users of any given platform should have a clear understanding of the behaviour that is or isn't acceptable on that platform. Governments and regulators should have a clear sense of what their powers are and the limits of those powers.

Clarity over objectives is essential for there to be trust that the framework is working. Without this clarity, stakeholders with competing interests may assume contradictory objectives and each argue that they are not being met. Policy-makers designing national approaches need to tightly define their objectives; if, for example, policy-makers' approach aims to protect against harm, this should include clear wording around the type of harm – for instance, physical, emotional, or psychological.

Greater clarity of objectives should also be a focus of international approaches. The G7's Internet Safety Principles set out broad wording around "Internet safety" and "the

[10] https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/the-limitations-of-automated-tools-in-content-moderation/.
[11] https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital; https://egazette.nic.in/WriteReadData/2021/225464.pdf.
[12] https://www.facebook.com/zuck/posts/10107243286682221.

exercise of all human rights online". Policy-makers should aim to build on this through iterative international discussions, for example by developing metrics around users' exposure to different types of illegal content and agreeing realistic and tangible targets for reducing that exposure.

Clarity over responsibilities is also important, particularly for avoiding a disproportionate impact on freedom of expression and a disproportionate commercial impact on intermediaries. A lack of legal certainty over the risk of enforcement action could chill innovation and competition by discouraging new, smaller companies from investing in intermediary services. Lack of clarity could also lead to platforms over-removing content to protect themselves from enforcement action.

As one example of the importance of clarity for establishing trust, there should be clear guidance for intermediaries on how to provide an effective framework for users to report illegal content. The expectations on intermediaries should also be communicated to users. This will help strengthen users' trust that, when they report illegal content, there is a clear process for what should happen next and how the intermediary is expected to respond. Policy-makers evaluating the success of their approaches should survey attitudes among different groups of users to assess the level of clarity they have on their own responsibilities as users and on the responsibilities of intermediaries.

2) **Adaptability** – Approaches should be designed so that they can adapt to the fast-changing nature of internet intermediaries themselves and the policy issues surrounding them.

The functionality of internet intermediaries is evolving all the time, challenging policy-makers to adapt their approaches. As one example, users' increasing awareness of the importance of privacy and the shift towards end-to-end encrypted private messaging tools has provoked debate on how to strike the right balance in protecting users in private spaces. Adaptability means having structures in place – at both the national and international level – for governments, intermediaries, user groups and rights experts to convene to discuss emerging issues and to agree shared approaches for addressing them. In the case of encryption, this should include evidence-based and inclusive dialogue on the risks and opportunities this creates for different groups of users, and on how those risks can be effectively mitigated. Accountability (Principle 3) – ensuring that what is agreed is implemented - and Cooperation (Principle 6) are also vital to this process.

There is the potential for tension between Adaptability and Clarity (Principle 1). There is an ongoing debate about whether responsibilities should be set through specific rules or through more general duties; for example, the UK's draft Online Safety Bill imposes a duty on intermediaries to take proportionate steps to mitigate and effectively manage the risks of harm arising from illegal content.[13] An approach purely focused on rules would provide maximum clarity, but could be too slow in the face of the evolving and diverse nature of online risks. Duties, by contrast, are more future-proofed, but if too broadly defined, may lead to platforms over-removing content to shield themselves from liability. Policy-makers should consider developing clear and tightly defined duties,

[13] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safety_Bill_Bookmarked.pdf.

retaining the flexibility to supplement these with more specific rules or guidance if they judge that greater clarity is required, for example if they assess that platforms are responding too slowly to a specific type of egregious illegal content.

Adaptability is vital for maintaining users' trust in the effectiveness of the framework. Measures of success could include: firstly, increased levels of confidence among users that when a new issue arises, there is a rapid and robust response from intermediaries and other stakeholders; secondly, using case studies to assess whether there has been such a response. For instance, policy-makers could evaluate the success of initiatives such as the Technology Coalition of intermediary companies, who have collaborated to address emerging problems such as self-generated indecent imagery featuring youth.[14]

3) **Accountability** – Approaches should strengthen accountability for governments, internet intermediaries and users, as well as accountability for how the system is working as a whole.

International and local approaches should provide accountability for intermediaries, by putting in place clear structures of oversight. This could include either an independent regulator or a co-regulatory body with powers to assess whether intermediaries are upholding their responsibilities and impose proportionate penalties when they are not. Internet intermediaries should also be accountable to their users, providing clear explanations in relation to content removal and recommendation. Intermediary accountability must be meaningful, while also taking account of the commercial impact; it may not be feasible, for example, for intermediaries operating at a global scale to provide detailed and bespoke explanations for each instance of content removal.

There should also clear accountability for users, including mechanisms for ensuring that users who break the law in posting content can be prosecuted. A key question relates to whether it should be possible for law enforcement to identify users' identity online in order to prosecute them for illegal activity. India's recent legislative proposal has been criticised for requiring intermediaries to provide the option for users to voluntarily verify their identities.[15] However, it is worth democratic countries exploring how, with robust safeguards such as independent judicial oversight, law enforcement should be empowered to identify individuals posting egregious, illegal content.

Any framework for intermediary governance itself also needs to be accountable. On a national level, to ensure that the framework continues to meet the evolving priorities and attitudes of a democratic country, this could include the oversight body having to account for its activity to the national parliament. On an international level, this should include the use of multilateral fora such as the G7 and OECD to develop KPIs to assess the effectiveness of international approaches such as the Internet Safety Principles in meeting their objectives.

Accountability is important for strengthening trust. Ultimately, the issues around harm and free expression online are too important to leave to private companies to decide alone. Democratic governments that are accountable to their voters need to assume a

[14] https://www.technologycoalition.org/category/news-announcements/.
[15] https://blog.mozilla.org/netpolicy/2021/03/02/indias-new-intermediary-liability-and-digital-media-regulations-will-harm-the-open-internet/.

leading role in outlining objectives and responsibilities for intermediaries, and in ensuring those responsibilities are upheld. Online user research often identifies the perception that intermediaries are not upholding their responsibilities; for example, a US survey found that 73% of US adults lacked confidence in technology companies' to prevent misuse of their platforms ahead of the 2020 Presidential Election.[16] The perception that intermediaries and other stakeholders understand there will be consequences if they do not uphold their responsibilities is vital for strengthening users' trust in the online environment.

4) **Empowerment** – Approaches should aim to increase individual users' choice and control in relation to internet intermediaries.

Intermediary approaches should, firstly, consider how to increase users' ability to choose within platforms, and secondly, consider how to increase users' ability to choose between platforms.

Within platforms, the evolution of the online communications environment is providing more opportunities to empower users, with an increased emphasis on customised and tailored experiences. Many platforms have developed tools enabling users to customise how, for example, advertising is targeted.[17] Intermediary approaches should further incentivise platforms to provide the ability for users to influence how content and advertising is presented to them, while being mindful of the commercial impact. For example, requiring detailed and lengthy opt-ins for targeted advertising risks negatively impacting not just on intermediaries themselves, but also on advertisers and publishers.

It is also important that users have the ability to decide between different services. This can help each user to choose a service in line with that user's preferences. Targeted, pro-competitive interventions, such as enabling users to port their data and contacts from one platform to another, could make it easier to switch platform. Policy-makers should again carefully consider the potential commercial impacts of these interventions. A European Commission report noted, for example, that full protocol interoperability – the development of standards allowing different services to interoperate – may reduce the ability of each service to innovate and differentiate from each other.[18]

Empowerment should also be a focus of international approaches, since it may help mitigate against the potential for cultural differences over the acceptability of different types of content. Governments from different parts of the world may be more likely to agree on the principle of empowering users to make their own decisions, rather than on whether content such as blasphemy should be allowed.

The difference in how individuals and societies perceive the balance of harm and free expression means that empowerment is important for achieving trust. Different groups of users also have different attitudes to different types of content, for example swearing, and providing choice can help all users to maximise the benefits of being online. To measure success, policy-makers should survey the level of confidence among different

[16] https://www.pewresearch.org/fact-tank/2020/10/27/how-americans-see-u-s-tech-companies-as-government-scrutiny-increases/.
[17] https://blog.google/products/ads-commerce/improving-user-privacy-in-digital-advertising.
[18] https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf.

groups of users that they have the ability to shape an online environment that works for them.

5) **Practicality** – Approaches should take account of the commercial impact and should preserve flexibility for continuous innovation by internet intermediaries.

Intermediary approaches need to set proportionate responsibilities of platforms. The oversight body should be required to take account of and minimise the risk that these responsibilities, firstly, lead to the over-removal of content, negatively impacting on users' freedom of expression; secondly, increase the burden on platforms to the point that it harms innovation and negatively affects smaller intermediaries' ability to compete with incumbents.

As one example, internet intermediaries operate at a scale that makes manual pre-checking of every individual piece of content all but impossible. According to one media agency, there are 243,000 photo uploads to Facebook every minute.[19] The automated tools used by the largest intermediaries do not have a 100% accuracy rate in determining whether content is illegal, and are more likely to incorrectly assess a piece of content as illegal in cases where context is more important, such as hate speech.[20] Smaller, less well-resourced intermediaries will have even less ability to manually pre-check content. This means that any approach that makes intermediaries directly liable for individual pieces of illegal content risks disproportionately impacting on freedom of expression and on competition and choice for users.

A more practical way of making intermediaries accountable would therefore be to set responsibilities focused on the systems and processes they have in place. The oversight body would then need to work with intermediaries to develop metrics for evaluating these systems' effectiveness in upholding their responsibilities. The oversight body would also need to develop a clear and proportionate framework for requesting information, to manage the risk of intermediaries being overwhelmed with urgent requests for data. As one example, Australia's draft Online Safety Act specifies that intermediaries must comply with requests for information from the eSafety Commissioner within 30 days.[21]

This emphasis on practicality will reinforce trust by addressing the concerns expressed by some citizens and digital rights groups that intermediary regulation risks leading to governments limiting free expression online. As part of their evaluation, policy-makers should gather evidence of compliance costs and undertake assessments of inadvertent impacts on freedom of expression, innovation and competition, to contribute to a wider dialogue on the framework's effectiveness in balancing its objectives.

6) **Cooperation** – Approaches should provide for transparent, inclusive, and effective cooperation between relevant stakeholders.

Accountability (Principle 3) needs to be supplemented with the ability for different stakeholders to cooperate and address tensions that arise. This is important, firstly,

[19] https://www.omnicoreagency.com/facebook-statistics/.
[20] https://journals.sagepub.com/doi/full/10.1177/2053951719897945.
[21] https://parlinfo.aph.gov.au/parlInfo/download/legislation/bills/r6680_first-reps/toc_pdf/21022b01.pdf;fileType%3Dapplication%2Fpdf.

within countries. Under any national framework, there should be mechanisms in place for the government, oversight body, industry, civil society and users to come together to address issues that emerge in relation to online intermediaries. These mechanisms should be inclusive, for example providing opportunities for children and vulnerable adult users to contribute.

Cooperation will also be important internationally to mitigate against intermediaries having to face competing and contradictory rules across jurisdictions. This could significantly increase compliance costs, negatively impacting innovation and the ability of new intermediary services to compete with incumbents, while risking inconsistent levels of protection for users in different countries. There are some important precedents for this cooperation, such as the Christchurch Call to Action, which sets out commitments from governments, industry and civil society in relation to addressing the risk of online terrorist content while safeguarding free expression.[22]

International cooperation will inevitably be challenging, as different countries seek to impose their own view of which content should and should not be allowed online. Within the EU, Hungary has expressed concern about Facebook limiting the visibility of "Christian, conservative right-wing opinions" and has passed its own law prohibiting displays of homosexuality to children.[23] Policy-makers in other EU countries, for example German MEP Tiemo Wölken, are concerned that Poland and Hungary will use EU-level legislation to "order the takedown of LGBT-related content".[24]

This underlines the importance that Cooperation be supplemented with Accountability (Principle 3) for national governments. Countries should use international fora such as the G7 and OECD to scrutinise whether national approaches are meeting basic standards set out in human rights law and in shared statements of policy such as the Internet Safety Principles. Realistically, national governments are likely to be pragmatic about how far they pursue accountability for each other. Nevertheless, policy-makers seeking to design effective international approaches should continue to aspire towards greater cooperation and accountability for governments.

Cooperation is essential to trust, because the inevitable differences in opinion about online content issues means that each stakeholder needs to feel that their voice is being listened to. It will not remove disagreement entirely, but policy-makers should at least aim to strengthen trust among both individual users and governments that there are fair and meaningful national and international processes for airing and resolving those differences.

[22] https://www.christchurchcall.com/christchurch-call-community-consultation-report.pdf.
[23] https://www.euractiv.com/section/digital/news/hungary-raises-concerns-about-shadow-banning-of-online-speech/; https://www.euractiv.com/section/non-discrimination/news/portrayal-and-promotion-hungarys-latest-anti-lgbt-law-explained/.
[24] https://calendar.boell.de/en/discrimination-by-moderation.