Digital
Action

><

Digital Action’s tech accountability policy taxonomy

This reference document outlines the types of government or company policy that can be used to limit the impact of online harms.

We have included those with the greatest **potential to have a positive impact on democracy, human rights and online harms, and that deal with the systemic problems with thead-tech business model.** They are:

1. Transparency
2. (Political) Advertising
3. Algorithms & Content Curation
4. Community Standards, Content Moderation and Company Enforcement
5. Liability & Enforcement
6. Privacy & Data Protection

For each policy area you will find a short explanation of why it is important in preventing online harms from impacting democracy and human rights. You will also find an outline of ‘what good looks like’ in each area and key considerations and examples for regulatory and non-regulatory approaches. There are also additional sources for further reading to guide decisions on the kinds of interventions we should support.

Policy Area	**Transparency**
Intro	Transparency spans across all other areas of policy and can refer to either transparency from governments (inc. regulators), or transparency from companies. It can be broken down into three key categories where transparency is required: policies (or laws), processes (or practises) and outcomes. It is important for a variety of stakeholders to fulfil their roles in online governance. Different levels of disclosure may be required for different stakeholders, for example to protect user privacy or Intellectual Property (IP): - Governments and elected officials require more information from companies to fulfil their obligations to citizens, including upholding and enforcing existing laws, providing democratic oversight, protecting national security, representing the victims of online harms, and advocating for change on their behalf. - Regulators also need a more complete understanding of company policies, procedures and decisions, as well the underlying technology, its outputs and potential biases. - Civil society, academia and the media require greater access to data to fulfil their public interest mandates and provide civic oversight - Transparency is vital for the public to understand their rights and responsibilities and how laws apply online, the relationships they enter into with online platforms, and the

	environment in which they spend increasing amounts of time, receive their information, and participate in society. Overall, transparency must complement rights to data privacy, not erode them. A good model for transparency will protect individuals' data privacy while enabling a macro understanding of the nature and scale of technology platforms' processes and any potential impacts on democracy, infringement of rights or harms that stem from their platforms. However, the requirements and expectations associated with transparency are often poorly articulated, lack specificity, or vary across online platforms and offline jurisdictions, so calls for further transparency should have a strong rationale and be as specific and targeted as possible.				
Democracy Features	**Rule of Law / Separation of Powers / Independent Judiciary**	**Transparency / Accountability in public admin.**	**Free & Fair Elections**	**Pluralistic political system**	**Free / plural media**
Impact / Why Important?	Transparency is vital to ensure and scrutinise government and/or law enforcement requests to and interactions with companies. Such requests must follow the rule of law, ensuring any legal processes are fair and contain the necessary safeguards. Transparency from companies is vital in terms of information they share with governments and/or law enforcement, and whether this follows due process and the rule of law.	Outside of formal legal processes, transparency is vital to understand informal government or state (including regulators) interactions with companies, for example where pressure is applied behind-closed-doors to companies to encourage them to make changes to their policies, processes or practices Transparency is a vital tool for regulators to be able to effectively oversee companies, but they must also be transparent themselves in their assessments, decisions and any sanctions or enforcement.	Transparency from governments or regulators is essential on efforts to combat election-related online harms during elections, to ensure that they do not disproportionately or unjustifiably disadvantage one side over another. Similarly, transparency from companies is essential on online political ads and election related online harms to ensure fairness and prevent abuses. Finally, transparency from political parties and election bodies is vital to understand how political advertising and online spending impacts elections.	Transparency is vital from both governments and companies to enable political opposition, civil society, activists etc. to scrutinise the prevalence of online harms, the impacts these can have on democracy, and the effectiveness of government or company actions to address these.	Transparency is vital from both governments and companies to enable media to scrutinise the prevalence of online harms, the impacts these can have on democracy, and the effectiveness of government or company actions to address these.

Human Rights	**Non-discrimination, Minority protections, Incitement, Security**	**Freedom of Expression / Belief / Association / Political Participation**	**Privacy / Defamation**	**Freedom of information**	**Necessary conditions & limitations**
Impact / Why Important?	Transparency is vital to understand the prevalence of discrimination or incitement online, whether in terms of policies, processes, or outcomes, whether by governments or companies, and including through the application of AI systems (see also Broader AI Concerns section in Annex)	Transparency is vital to understand infringements on freedoms of expression, belief, association, participation online, whether by companies or states, across policies, processes, and outcomes.	Transparency is vital to scrutinise and ensure any government or law enforcement data requests to companies to infringe on privacy are legal (limited, proportionate etc.) Companies must be transparent in terms of how personal data is collected, processed and used, and who it's shared with.	Transparency from companies on how online content is amplified, recommended, moderated or removed is vital to understand how online platforms operate and how they impact the flow of information across the online ecosystem.	See Freedom of Expression etc. – transparency is required to judge the necessity and proportionality of any restrictions on or infringements of rights by governments or companies.

Online Harms	**Disinformation & Misinformation**	**Hate Speech & Incitement**	**Online Abuse & Harassment**	**Online censorship**	**Invasions of Privacy**
Impact / Why Important?	Lack of transparency from platforms prevents effective third-party analysis and/or oversight of the extent and nature of online dis/misinformation, hate speech or abuse, and any company efforts to prevent or mitigate these harms, whether by governments, regulators or academics and civil society researchers.			Online censorship is often shrouded in a lack of transparency from both governments and tech companies around government takedown and/or data requests made to platforms, and behind-closed-doors pressure from governments on companies to moderate a wide range of illegal and/or legal (but harmful content), or censor dissenting opinions, leading to a lack of accountability and rule of law, due process and opportunities for redress. Online censorship is often heightened during sensitive or dangerous political moments (for those in power, e.g. elections, protests) and can provide cover for human rights abuses or abuses of state power (e.g. state violence, corruption), and prevent effective documentation of such abuses.	The weakness of data privacy laws and/or general lack of (genuine) transparency for the public on how tech companies collect, use and share their data (e.g. through inaccessible T&Cs) limits public understanding of, and meaningful consent to, these practices. Oversight by governments or regulators (if in place) is typically retrospective, with any sanctions or remedies enforced after-the-fact, undermining the extent to which accountability can lead to practical changes or limit negative impacts (e.g. when employed following an election).

What does good look like?	**Governments:** - Transparency of government requests to companies (both formal and informal). - Transparency of policy or regulation development processes. - Transparency of any regulatory assessments, investigations, oversight or actions and interactions with companies. - Transparency of any law enforcement or legal action. - Transparency of political advertising.	**Companies:** - Policies: transparency around the content of policies, and who (inside and outside the companies) is consulted in their development. - Processes: transparency around how company products are developed, how platforms and their features (including algorithms) work, and how policies are applied in practice (including resourcing, overall responsibility etc.). - Outcomes: transparency around harms, responses (i.e. moderation, redress), and impacts on democracy and/or human rights.
Regulation *See 'Levers' section of framework for key considerations checklist for non-regulation*	**Key considerations:** - Transparency is often among the least controversial policy areas, and one area that opposing sides of policy debates (e.g. freedom of expression vs. harms focused) can agree on as foundational to any regulation, but may still be resisted by companies reluctant to reveal more. In contexts where there is little agreement over whether or how to introduce additional regulation, transparency often represents a 'safe bet' as a starting point that different stakeholders can agree to, and may provide the basis for further regulation later on. - Transparency should cover both companies and governments/regulators themselves, and cover all aspects of their operations (i.e. policies, processes, outcomes). - There may need to be different levels of transparency from companies for different stakeholders – i.e. regulators, law enforcement, academia, civil society, the general public etc. For example, the general public may not require access to internal company documents, an API or IP on algorithms, but these should be clearly explained at a non-technical level to build understandings of their basic workings and allow for greater scrutiny of outcomes, whereas regulators will require greater transparency and access to understand how AI is designed, what it is optimised for etc. - Transparency needs to be regular and ongoing to keep pace with the evolving nature of online platforms, harms, crises etc., rather than for example annual statements that provide aggregated data in a manner that prevents effective scrutiny. - Transparency should not necessarily be uniform across platforms to account for the wide variety of different platform functions and designs. Similarly, different online harms will require	**Examples** **EU Digital Services Act** Overview of Transparency Obligations for Digital Services in the DSA (CDT) **UK Online Safety Bill** The government report on transparency reporting in relation to online harms (DCMS, Consultation outcome, Dec 2020) Transparency in the regulation of online safety (Ofcom, May 2021) **German NetzDG** Company NetzDG Transparency Reports: FB, G, TW, Regulating transparency?: Facebook, Twitter and the German Network Enforcement Act (Conference on Fairness, Accountability, and Transparency, January 2020)

	different types of data. However, some level of consistency in definitions, metrics and baselines is required to allow for comparisons across platforms. Regulation should recognise the varying size, resources and risks of different companies and platforms, and ensure that transparency requirements are proportionate to avoid overburdening smaller or less risky or impactful platforms.	
Non or Self-regulation *See 'Levers' section of framework for key considerations checklist for non-regulation*	**Key considerations:** - Incremental progress has been achieved in many areas over the past 5-10 years through government, civic and media pressure on companies to be more transparent about their policies, processes, and outcomes. However, without mandating transparency via regulation, companies are still relatively free to act as gatekeepers and determine what information and data to release, in what form (i.e. metrics), how often, and determine how it is presented. - Fundamentally, ad-tech companies have a vested interest in less transparency to protect their business models. They lack the commercial and PR incentives to be more transparent, as there is a danger that this would reveal the true scale of problems or harms on their platforms, the relative ineffectiveness of responses (e.g. the extent of content moderation across languages), the role and impact of their own ad-tech systems and platform design features, commercial sensitivities around algorithms, and claims that too much transparency benefits bad actors (e.g. Facebook Is Failing in Global Disinformation Fight, Says Former Worker – NYT, Facebook tries to block tool aimed at promoting transparency around political ads - Politico). - As a result, non-regulatory attempts to push companies to be more transparent must be carefully coordinated, and ideally combine public and private pressure from a variety of different stakeholders in order to be effective. Most successful non-regulatory examples have either achieved narrow and specific changes on a company-by-company basis (often based on upholding existing commitments, or providing further clarity where metrics are deliberately opaque), or worked across industry and with international organisations to secure collective commitments from companies.	**Examples** **EU Code of Practice on Disinformation** EU Code of Practice on Disinformation: Briefing Note for the New European Commission (Carnegie) Cracking the Code: An Evaluation of the EU Code of Practice on Disinformation (ISD) **Global Internet Forum to Counter-Terrorism (GIFCT)** GIFCT Transparency Report Raises More Questions Than Answers (Brennan Center, Sept 2019) Human Rights NGOs in Coalition Letter to GIFCT (Various, July 2020) A Human Rights Assessment of the GIFCT (BSR, July 2021) **Christchurch Call** The Christchurch Call: The Good, the Not-So-Good, and the Ugly (EFF, May 2019) Christchurch Call emphasizes human rights, but needs meaningful participation and transparency for rights-respecting outcome (Access Now, May 2019) Why we need more transparency to combat terrorist and violent extremist content online (OECD, Sept 2020) **Tech Against Terrorism** The Terrorist Content Analytics Platform and Transparency By Design (Nov 2020) **Social Science One** Ahead of 2020, Facebook Falls Short on Plan to Share Data on Disinformation (NYT) The Social Science One Facebook Cooperation: A Systemic Failure

🟩		(Political Data Science) Creating a French framework to make social media platforms more accountable: Acting in France with a European vision (“Regulation of social networks – Facebook experiment”, Submitted to the French Secretary of State for Digital Affairs, May 2019)

Policy Area	**(Political) Advertising**				
Intro	Social media companies' ad-tech business models – i.e. the use of systems, tools and services that connect advertisers with target audiences of non-paying users to generate profits - rely on the collection and processing of vast amounts of personal data in order to target advertising at specific subsets of online users, and have become the driving force of the online economy. This collection of data is typically legal with the consent of the user, but is a condition of access to online platforms, services and/or content. This consent is often secured through processes marked by stark information and power differentials (e.g. lengthy and highly legalistic and inaccessible Terms of Service, unclear explanations of how data is used, and who it may be shared with) that call into question the possibility of *meaningful* consent. This is exacerbated by the dominance of the largest companies in many contexts or markets, where such platforms have become near-essential forms of communication (and also moved into other sectors, e.g. online payments). Additional data collection is also conducted by using tracking cookies that follow users across other websites, and shared logins that connect user's activity across different online services to build up a more comprehensive data-set of their online behaviours and interests. As well as targeting advertising, this data is used internally by companies to 'optimise' their platforms, both in terms of their design and the types of content that are 'served' or recommended to individual users via algorithmic systems (e.g. newsfeeds), with the aim of maximising the amount of time users spend on the platform with the ultimate goal of maximising potential advertising revenues. These same tools and systems, while fundamentally designed for commercial purposes (and to generate commercial rather than political revenues), have also been heavily adopted by political actors to target political messaging at highly specific groups in a way that has not been previously possible with other forms of offline advertising (e.g. broadcast, print, billboards etc.), which has had a significant impact on political communications and campaigning. Finally, there is a growing debate around the effectiveness (and therefore sustainability from an advertiser perspective) of individually tailored and targeted advertising based on personal data, when compared to contextual advertising with targeting based on the content which an advert will appear next to online. Similarly, legal action has been taken challenging the accuracy of the advertising metrics provided by ad-tech companies to their advertising customers (known as 'ad fraud'), and therefore the value for money provided to businesses from online advertising.				
Democracy Features	**Rule of Law / Separation of Powers / Independent Judiciary**	**Transparency / Accountability in public admin.**	**Free & Fair Elections**	**Pluralistic political system**	**Free / plural media**
Impact / Why Important?	The scale and lack of real-time transparency and oversight of political advertising in many contexts means that	Transparency and oversight is often lacking in terms of how governments and/or political parties and other	Ad tech business models, and the data collection and micro-targeting that underpins them, can	Both pros and cons: larger established political parties are able to maintain and exploit existing advantages in terms of resources, access to and processing of data etc., potentially entrenching their positions.	Data-driven ad tech business models have undermined traditional media business models (i.e. big tech dominating advertising revenues, resulting in cutting staff, local or investigative reporting for

	any investigations, redress or sanctions are necessarily retrospective (required to follow rule of law principles) – which can be especially problematic in for e.g. elections or incitement related contexts where impacts on democracy or human rights cannot easily be reversed or effectively compensated.	political actors use online advertising – existing regulation in these areas for offline advertising has often not been updated to keep pace with technological changes, leading to an accountability vacuum. For example, companies have been pressured (initially using non-regulatory levers, but increasingly through regulation) to create online archives of political advertising, but to date these have often been incomplete or missing key types of information.	enable targeted voter suppression (through online advertising, disinformation, threats etc.) of specific, minority or marginalised communities. Invasive data collection and ad-targeting can lead to the atomisation of political campaigning, enabling highly specific targeting of different messaging to different voters, with limited or no transparency to highlight contradictions in political parties' messaging or platforms.	On the other hand, smaller parties may be able to use online advertising to increase their reach and exposure more cost-effectively than they would be able to offline, and target their messaging more precisely than via other forms of media *(Note: with the caveats that a) this is often the companies' line so shouldn't necessarily be taken at face value, and b) in many contexts this has favoured more fringe or extreme parties, which may increase the pluralism of the political system, but also incentivise populism/polarisation etc.)* Overall ad-tech systems and prioritisation of engagement can help parties to generate unpaid visibility, but also incentivise more divisive, misleading or abusive content to drive engagement.	many outlets etc.) and changing editorial incentives (e.g. encouraging clickbait articles that are cheap to produce and drive ad revenues over in-depth, investigative or local reporting) – with the overall impact of undermining the independence, plurality and quality of media *(Note: with the caveat that online platforms have also provided opportunities for non-traditional outlets to grow their audiences, even if these are often more partisan/less balanced to thrive in online environment).*

Human Rights	**Non-discrimination, Minority protections, Incitement, Security**	**Freedom of Expression / Belief / Association / Political Participation**	**Privacy / Defamation**	**Freedom of information**	**Necessary conditions & limitations**
Impact / Why Important?	Online ads can be used by political parties to target highly specific or localised audiences, which can lead to discrimination against or neglect of certain groups, and overall leads to the atomisation of political debates. Online ads can be used to incite discrimination and/or violence against minority or marginalised communities, often with insufficient oversight from companies and/or regulators.	Pros / cons similar to impacts on plurality of political systems above – i.e. online advertising can provide benefits in terms of political participation by lowering the costs and improving the targeting of advertising for issue-based campaigns, but overall the impacts of the ad-tech business model in	Ad-tech business models are the primary driver behind extensive and invasive data collection by platforms that systemically undermine privacy, and create broader vulnerabilities to a range of online harms (e.g. targeted abuse	The ad-tech business model overall creates a distorted online information ecosystem or environment, creating a non-level playing field where certain types of content thrive (e.g. clickbait, inflammatory etc.). Online advertising is also used to artificially distort the information environment e.g.	Governments should not attempt to exert undue control over the distribution of legitimate, legal speech online (for example to disadvantage, censor or persecute political opponents) via regulation of social media algorithms. Governments may only constrain rights such as freedom of speech and freedom to information to protect the rights of others, and under specific exemptions such national security or public health. Any government intervention that impacts these rights must be legal,

	Online advertising has also enabled or entrenched broader forms of discrimination via advertising for jobs, housing etc. where non-discrimination legislation has not kept pace with technological changes.	catalysing a variety of online harms also play a significant role in undermining political participation or chilling freedom of expression (see Harms below).	such as doxing, electoral interference/disinformation) that impact other human rights and features of democracy.	oil companies and fossil fuel interests using online advertising to influence and distort online discussions around climate change.	necessary and proportionate (see Foundational Principles), and conducted by an independent regulator.

Online Harms	**Disinformation & Misinformation**	**Hate Speech & Incitement**	**Online Abuse & Harassment**	**Online censorship**	**Invasions of Privacy**
Impact / Why Important?	Ad tech business models, and the data collection and micro-targeting that underpins them, can enable targeted voter suppression (through online advertising, disinformation, threats etc.) of specific, minority or marginalised communities. Invasive data collection and ad-targeting can lead to the atomisation of political campaigning, enabling highly specific targeting of different messaging to different voters, with limited or no transparency to highlight contradictions in political parties' messaging or platforms, or identify for e.g. foreign interference in elections using advertising.	Ad-tech business models based on data collection and processing enables hate speech or incitement to be targeted at specific audiences or marginalised or minority communities. Hateful or discriminatory attacks on reputation, infringing privacy of individuals (e.g. hate speech combined with doxing). Prevalence and extent of online data (as a result of the ad-tech business model) enables surveillance or hacking by states or non-state groups, which can then be used as a basis for hateful or discriminatory attacks.	Ad-tech business models based on data collection and processing enables online harassment, abuse or incitement to be targeted at specific audiences. Online harassment or abuse targeting reputation, infringing privacy of individuals (e.g. defamation and doxing) Prevalence and extent of online data (as a result of ad-tech business model) enables surveillance or hacking by states or non-state groups, which can then be used as basis for online harassment or abuse	Ad-tech business models have increased the ability of governments to target opposition, activists, civil society etc. dissent through the abundance and availability of data online, with online censorship laws often accompanied by measures that infringe on privacy and/or anonymity online.	Ad-tech business model based on data collection and processing enables dis/misinformation to be targeted at specific audiences. Dis/misinformation used as a tool for attacks on reputation, infringing privacy of individuals. Prevalence of online data (as a result of ad-tech business model) enables surveillance or hacking by states or non-state groups, which can then be used as basis for dis/misinformation (including partial disclosure, doctored information etc.).
What does good look like?	**Governments:** Data and privacy regulation should include limits on what sensitive		**Companies:** Companies should provide accessible, timely and *computational* transparency (i.e. via well-functioning		

	personal data can be collected and combined, and consequently used for ad targeting purposes to discrimination, limit the potential impact of data breaches. Such regulation should incorporate effective and timely oversight and enforcement powers for data regulators, and should ensure or encourage companies to employ stronger safeguards to limit the collection, retention and sharing of user data, through for example "know your customer" requirements for social media platforms. Electoral laws should be updated to incorporate clear rules for online political advertising, including measures around transparency for advertising content (e.g. imprints), targeting, and spending. Again this should include sufficient powers for regulators to provide timely oversight, and have sufficient investigative powers and options for sanctions to disincentivise infringements. Competition regulation may also be required to counter the dominance of the (online) advertising market by a small number of large online platforms such as Facebook and Google (see Competition, Monopoly & Democracy section in Annex).	APIs and/or ad-libraries) on online advertising, with additional levels of transparency for political or issue-based advertising for citizens/users, regulators, and academics, researchers and journalists. This should include who advertisers are, how much they are spending, who they are targeting and how, and the full range of ads they have placed/have running. Companies should conduct regular, proactive and effective checks and enforcement on *advertisers*, especially new customers, for example through "know your customer" approaches, combined with sufficient penalties for advertisers that disincentivise infringements. Companies should employ more effective and sophisticated approaches to monitoring advertising *content*. For example, many ads run on platforms that contravene company advertising policies, and checks are often limited to partial or incomplete lists of prohibited keywords, or overly rely on automated systems with significant flaws. Sufficient investment and resourcing for human review is required to combat false negatives (i.e. content that shouldn't be allowed getting through) and false positives (i.e. civil society or campaigner ads being blocked incorrectly for addressing 'political' issues). Companies should develop more effective enforcement around the content that advertising appears next to prevent the monetisation of harmful content. For example, online advertising has created a viable business model for producers of harmful content on YouTube, and incentivised click-bait etc.
Regulation *See 'Levers' section of framework for key considerations checklist for regulation*	**Key considerations:** - One of the key challenges in regulating advertising is how to define 'political' or 'issue-based' advertising, e.g. where is the boundary between commercial and political intent (e.g. oil companies promoting supposedly green energy), and who decides? - Transparency is essential for effective oversight and enforcement, and must be as proactive as possible in order to be effective. - Mandated access for third parties to ad-libraries and archives – i.e. journalists, academics, civil society, public (e.g. info available on how/why individuals are targeted, see the ads they do). - Regulation of both platforms and advertisers should be in-line with or broadly comparable to other mediums used for advertising (e.g. broadcast, print).	**Examples** **EU Digital Services Act (EU Commission)** European Democracy: Commission sets out new laws on political advertising, electoral rights and party funding (EU Commission, Nov 2021) #WhoReallyTargetsYou: DSA and political microtargeting (EDRi) Europe offers tepid set of political ads transparency rules (TechCrunch, Nov 2021) **Elections Modernization Act & Registry Requirements for Political Ads on Online Platforms (Elections Canada)** What have we learned from Google's political ad pullout? (IRPP) **Proposal to Regulate Transparency of Online Political Advertising (Rep. of Ireland Government)**

		New election laws to include crackdown on political ads on social media (Irish Times, Jan 2021) Facebook urges Ireland to hold off on political ads rules until EU legislation (Euractiv, Mar 2021) U.S. charges Facebook with racial discrimination in targeted housing ads (Reuters, March 2019) Are Facebook Ads Discriminatory? It's Complicated (Wired, Nov 2019) Facebook still runs discriminatory ads, new report finds (The Verge, Aug 2020) Solving the problem of racially discriminatory advertising on Facebook (Brookings, Oct 2021)
Non or Self-regulation *See 'Levers' section of framework for key considerations checklist for non-regulation*	**Key considerations:** - It is inherently challenging to push for better non-regulatory approaches in this area as advertising is so central to company business models, and therefore they are likely to be reluctant to address concerns unless it becomes a business (i.e. advertisers' concerns over inaccurate metrics/ad-fraud) or a PR issue (i.e. negative media coverage impacting advertisers). - However, as in most contexts *political* advertising only makes up a small proportion of companies' ad-revenues, there may be scope to push for changes that have a positive impact on harms caused by political advertising, and limit companies' exposure to pressure from governments, regulators, advertisers or the media, but do not have a huge impact on them from a business or financial perspective. Ultimately, short of new regulation, advertisers themselves are often likely to be most influential in forcing changes from companies.	**Examples** **EU Code of Practice on Disinformation** EU Code of Practice on Disinformation: Briefing Note for the New European Commission (Carnegie) Cracking the Code: An Evaluation of the EU Code of Practice on Disinformation (ISD) Commission pushes for 'timely' update of disinformation code of practice (Euractiv) Platform ad archives: promises and pitfalls (Internet Policy Review, 2019) Ireland's Abortion Referendum Becomes a Test for Facebook and Google (NYT, May 2018) Foreign groups invade Ireland's online abortion debate (Politico, May 2018) How Ireland Beat Dark Ads (Foreign Policy, June 2018) Facebook to publish data on Irish abortion referendum ads (Guardian, July 2018) **Voluntary pledges for political parties:** First national Code of Conduct on online political advertising in the European Union signed by Dutch political parties and global online

🟩		platforms (IDEA, 2021) The Pledge for Election Integrity (Alliance of Democracies - various/candidate focused) The Coalition for Reform in Political Advertising (UK, 2019 - unsuccessful)

Policy Area	Algorithms & Content Curation				
Intro	The majority of social media platforms, and all the major ad-tech platforms, use algorithmic systems to determine which users see which content, and how often. These systems combine individual user data (including demographics, interests and behaviours) with data on the nature of content (origin, topic, how engaging it is etc.) to provide personalised content to each individual user to maximise the time that user spends on the platform, typically via 'newsfeeds' or recommendations. Algorithms are also used to increase the density of networks on a platform by providing suggested new connections, whether other users that individual may know, or pages, channels, groups or prominent accounts that may be of interest based on their personal data. These algorithms are constantly updated and optimised via feedback loops to ensure that they incorporate new data and signals indicating whether previous recommendations were effective in generating engagement from each user to maximise their value to advertisers. Ad-tech companies often claim neutrality, reject more traditional labels such as 'publisher', and are typically not liable for the content on their platforms (see Liability section below). However, the way in which they have designed their platforms and the online spaces they contain can have a significant impact on how users interact and what types of content are most popular and/or engaging. Recent years have seen a proliferation of discussion over the impacts of these systems and their impacts on democracy, from 'echo chambers' to 'filter bubbles' which, while contested, have become increasingly prominent in policy debates. While many platforms have been pushed to provide users (and other stakeholders, from governments and regulators to civil society, academia and journalists) with more transparency and choice over how these systems operate and influence their online experience, they are also reluctant to allow more substantial or independent transparency as these systems are so central to their business models and are closely guarded commercial IP.				
Democracy Features	**Rule of Law / Separation of Powers / Independent Judiciary**	**Transparency / Accountability in public admin.**	**Free & Fair Elections**	**Pluralistic political system**	**Free / plural media**
Impact / Why Important?	Concerns of too much state control over speech in both democratic and non-democratic contexts. Any regulatory oversight of platform algorithms must be independent of government and/or political interference to ensure platforms are not intentionally biased in favour of any particular	The current lack of transparency in terms of how algorithms are designed, developed and optimised leads to a lack of understanding and therefore accountability in terms of their impact on democracy (via for example debate over filter bubbles or echo	Platform algorithms and content curation can cause distortions of the information ecosystem, and while evidence in terms of platforms' political bias (e.g. favouring left or right) is not conclusive across different contexts, the impact that these systems have on the types of content that are most successful on ad-tech platforms may have an impact by for example amplifying the most polarising, inflammatory or even inciting content. This may be particularly impactful during		Platform algorithms can have significant impacts on broader media ecosystems, where media outlets have to optimise content to be successful on dominant ad-tech driven platforms. This can have an impact on incentives for media, for example by reward click-bait or inflammatory reporting, or creating additional pressures to favour speed over accuracy in reporting. Platform algorithms, and the overall ad-tech business model has also had a significant impact on the traditional business models of media as they have been squeezed out of the advertising market. This can disincentivise investment in investigative and/or local reporting, and leaves media outlets vulnerable to sudden changes by platforms (e.g.

	political actors (i.e. current governments).	chambers driving polarisation).	key political moments such as elections, but may also play an ongoing role in driving partisanship or division in the longer-term.	*How Facebook's Chaotic Push Into Video Cost Hundreds of Journalists Their Jobs*, *The Atlantic*)

Human Rights	**Non-discrimination, Minority protections, Incitement, Security**	**Freedom of Expression / Belief / Association / Political Participation**	**Privacy / Defamation**	**Freedom of information**	**Necessary conditions & limitations**
Impact / Why Important?	Algorithms can significantly amplify harmful and/or discriminatory content at speed and scale unmatched by company moderation efforts, potentially leading to offline violence in the worst cases (for e.g. Myanmar). Algorithms, and the data harvesting systems than underpin them, play a key role in targeting advertising and therefore contribute to various other forms of discrimination (see Advertising section above).	Algorithms can have a significant impact on what content is or isn't amplified (and therefore most successful or engaging) on dominant platforms, which are therefore not neutral and create an uneven environment for expression and/or political participation. This amplification can exacerbate various forms of online harm (e.g. dis/misinformation, hate, abuse etc.) that have a chilling effect on freedom of expression, belief or association, and serve to reduce political participation, especially among minority or marginalised communities.	Ad-tech business models and resultant engagement-driven algorithms optimise time spent on platforms via collection of vast amounts of granular user data on interests, behaviours etc., at the expense of user privacy and typically without genuine informed consent of users. Content algorithms can also play a role in amplifying false, defamatory, or non-consensual private content (e.g. images, addresses etc.), thereby exacerbating various forms of online harm, from dis/misinformation to abuse and harassment.	Algorithms determine and distort the flow of content and information across social media platforms, and therefore do not provide a wholly unmediated or free choice to information consumers. *(Note: important caveat that this has always been the case pre-internet in other mediums too, e.g. print, broadcast etc. are mediated too, and the application of this right to social media platforms and content curation is relatively new and untested)*	Governments should not attempt to exert undue control over the distribution of legitimate, legal speech online (for example to disadvantage, censor or persecute political opponents) via regulation of social media algorithms. Governments may only constrain rights such as freedom of speech and freedom to information to protect the rights of others, and under specific exemptions such national security or public health. Any government intervention that impacts these rights must be legal, necessary and proportionate (see Foundational Principles), and conducted by an independent regulator.

Online Harms	**Disinformation & Misinformation**	**Hate Speech & Incitement**	**Online Abuse & Harassment**	**Online censorship**	**Invasions of Privacy**
Impact /	The design of ad-tech platforms' content curation algorithms to			It is often unclear what is de- or not prioritised by	Ad-tech business models are built on a

<table>
<tr>
<td><b>Why Important?</b></td>
<td>maximise engagement can lead to the amplification of a variety of online harms, increasing their speed, reach and impact.

These systems can also have a broader impact by incentivising false, sensational, polarising or harmful, abusive or inciting content in online spaces. Content curation algorithms can be exploited or ‘gamed’ by bad actors to increase the reach and engagement of harmful material, as well as grow their networks and increase their audiences (and resulting advertising incomes).

Currently there is little transparency or accountability for ad-tech algorithms’ impact on these online harms, from their design and purpose, through to their outcomes and impacts.</td>
<td colspan="2">content curation algorithms, who decides, and why, leading to accusations that platforms have a political bias, and that practices such as “shadow-banning” amount to online censorship (i.e. limiting the exposure of certain content or accounts by not including in algorithmic recommendations or news feeds). In some instances these accusations of bias may be unproven, or the complaints of those whose content has been moderated to reduce harms (or may not have been successful in the first place). However, biased algorithmic systems can also negatively impact those trying to raise awareness of offline abuses or offline violence, disproportionately impacting already marginalised or minority communities.</td>
<td>combination of extensive data collection and content curation algorithms designed to prioritise engagement, ultimately to maximise advertising revenues.

This invasive data collection increases users’ vulnerabilities to invasions of privacy and other online harms by building highly sensitive ‘profiles’ of each user.</td>
</tr>
<tr>
<td><b>What does good look like?</b></td>
<td colspan="2"><b>Governments:</b>

- Any regulation that impacts social media platforms’ content curation systems must be independent from government and any political interference, and meet key criteria to comply with human rights (see above).
- Governments should focus on the overall impact of these systems on democracy, human rights, and online harms through risk-based ‘safety-by-design’ approaches. These should include effective due diligence from companies, and a regulator with sufficient resources and expertise to provide effective oversight.
- Any regulation should start with a focus on improving transparency to better understand the impacts of social media platforms’ content and recommendation systems, including data access for regulators and third-party experts.
- Governments should consider content curation and recommendation algorithms as constituent parts of the wider ad-tech business model, and take holistic approaches that recognise their interconnectedness with online advertising, data collection etc.</td>
<td colspan="2"><b>Companies:</b>

- Companies should be transparent and accountable for their algorithmic policies, processes and outcomes. This should include information on: design and optimisation decisions; training data; risk/bias assessments and internal research; evidence of outcomes and trade-offs from the application of algorithms; and the effectiveness of risk mitigations.
- Internally, companies should ensure those involved in AI design and implementation are more diverse, conduct sufficient risk and assessments, and employ safety-by-design principles to better consider implications of design/optimisation choices, and pre-empt and mitigate harms.
- Companies should invest sufficient resources to ensure equitable outcomes across all contexts, for example by not prioritising certain markets over others.
- Companies should provide users with more accessible options to customise the types of content they receive in their newsfeeds or recommendations, along with more public education on the use of algorithmic systems.</td>
</tr>
<tr>
<td><b>Regulation</b></td>
<td colspan="2"><b>Key considerations:</b></td>
<td colspan="2"><b>Examples</b></td>
</tr>
</table>

See ‘Levers’ section of framework for key considerations checklist for regulation	- Governments and regulators should learn from approaches to algorithmic oversight in other fields and sectors, and ensure they collaborate effectively with other governments and regulators internationally. - Given the potential dangers to freedom of expression and information of government regulation of communications, effective and proportionate regulation requires ongoing research and transparency around the negative impacts of content curation algorithms. This requires sufficient access, expertise and resources for both regulators and third-party experts (academia, civil society etc.). Any interventions or changes required by a regulator should be carefully tested and considered, with an awareness that small changes can have big effects given the scale of major ad-tech social media platforms. - While social media content curation systems are one application of algorithms that play a key role in shaping the online environment, they are also used increasingly in many other areas (including content moderation, see below). Governments should therefore also consider the broader impacts of AI, and consider whether additional regulation is required (see Broader AI Concerns in Annex).	**EU Digital Services Act** Panoptykon (Polish NGO) Can the EU Digital Services Act contest the power of Big Tech’s algorithms? (EDRi, Aug 2021) EU: Regulation of recommender systems in the Digital Services Act (Article19, May 2021) Make online platforms accountable for their algorithms, leading MEP says (Euractiv, Nov 2021) **UK Online Safety Bill** Examining the Black Box: Tools for assessing algorithmic systems & Technical methods for regulatory inspection of algorithmic systems (Ada Lovelace) Algorithm Inspection & Regulatory Access (DIgital Action / various NGOs)
Non or Self-regulation *See ‘Levers’ section of framework for key considerations checklist for non-regulation*	**Key considerations:** - Companies are careful to position their content curation systems as politically neutral and downplay their role in amplification of harmful content to avoid potential liability stave off regulation. Political, civic and media pressure has been successful at prompting changes by companies to their algorithms in some limited areas over time, but more fundamental changes have been difficult to secure, and algorithmic systems have rarely been addressed through formal self-regulatory instruments, as these systems are so central to ad-tech business models. - There is still very little transparency from companies when not compelled by regulation, and they often claim that privacy and commercial sensitives prevent further transparency on the impact of these systems on democracy, human rights, and online harms across different contexts.	**Examples** **Facebook & Instagram:** Why Facebook's news feed is changing – and how it will affect you (Guardian, Jan 2018) Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead. Internal memos show how a big 2018 change rewarded outrage and that CEO Mark Zuckerberg resisted proposed fixes (WSJ, Sept 2019) Facebook reportedly ignored its own research showing algorithms divided users (The Verge, May 2020) How Facebook shapes your feed: The evolution of what posts get top billing on users’ news feeds, and what gets obscured (WSJ, Oct 2021) Facebook’s race-blind practices around hate speech came at the expense of Black users, new documents show: Researchers proposed a fix to the biased algorithm, but one internal document predicted pushback from ‘conservative partners’ (WSJ, Nov 2021) Facebook tests giving more control of News Feed content to users again (The Verge, Nov

	They can be selectively transparent or point to contradictory research, but at times have acknowledged the role their systems can play in amplifying harmful content.	2021) How Instagram's algorithm is censoring women and vulnerable users but helping online abusers. (Are, C., Feminist Media Studies, 2020) **YouTube:** YouTube introducing changes to give people more control over recommended videos (The Verge, June 2019) YouTube's recommender AI still a horror show, finds major crowdsourced study (TechCrunch, July 2021) How TikTok Reads Your Mind: It's the most successful video app in the world. Our columnist has obtained an internal company document that offers a new level of detail about how the algorithm works (NYT, Dec 2021) The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems (Access Now, Amnesty, HRW, Wikimedia)

Policy Area	**Community Standards, Content Moderation and Company Enforcement**
Intro	All the major ad-tech platforms have developed their own unique sets of rules around what content and behaviours they allow on their platforms, typically spanning a wide range of online harms, and often known as ‘Community Standards’ or ‘Community Guidelines’ (or similar, see here). These standards cover content or activities that are illegal in most jurisdictions (e.g. ranging from terrorist or child-sexual abuse content to copyright infringements), but also a wide range of other types of content or activities that are not necessarily illegal in many contexts, but that companies may wish to prohibit (e.g. ranging from dis/misinformation, self-harm, to nudity or sexual content, spam etc.). There are also types of content that may be illegal in certain countries, but that companies do not prohibit outside of those particular jurisdictions, if at all (e.g. blasphemy, or previously Holocaust denial). This combination of illegal and legal (but possibly harmful) content, jurisdictions, and differing rules across platforms leads to a complex patchwork of rules online, with significant ‘grey areas’ at the margins. This is further complicated by the fact that companies typically have one set of rules for the public, but then more detailed internal guidance on how these rules are (meant to be) applied in practice through company moderation and enforcement. Additionally, both these internal and external sets of rules are constantly evolving as platforms grow and react to external pressure (see the Platform Governance Archive which tracks these changes over time for key platforms). If these rules or ‘Community Standards’ represent companies’ *policies*, then their practical application is dependent on internal company *processes* for content moderation, and company enforcement decisions and actions representing the *outcomes* of these processes. Company processes for content moderation include user-facing options to ‘flag’ or report content directly to the company (including ‘trusted flagger’ programmes that prioritise reports from expert partners e.g. fact-checkers, NGOs), and some limited proactive enforcement by specialist teams within companies that focus on particular types of infringing content or behaviour (e.g. terrorism, state-backed disinformation campaigns). However, they are increasingly reliant on (outsourced) human moderators and automated systems that use algorithms to identify and make decisions on potentially violating content at scale. In terms of enforcement, companies typically employ an escalating scale of responses to prohibited content or behaviour, ranging from warnings and/or limiting reach or engagement (e.g. removing from algorithmic recommendations, disabling comments etc.), through to content removal and eventually account suspension or removal. The large ad-tech platforms do have appeal processes for content moderation, but these are also inconsistent and opaque, with only a minority of cases ever going through courts. In the most serious cases concerning illegal content (e.g. terrorism, child-abuse), companies may also collaborate with law enforcement, however this varies considerably across different categories of content and jurisdictions. Across all three areas, companies are regularly criticised for their constantly evolving policies and processes, and lack of consistency in terms of outcomes. This can manifest very differently across different areas of policy and geographic contexts, with companies criticised for both under and over-moderation and enforcement and the resultant impacts on democracy and human rights. This can either be as a result of unclear or ill-thought out policies, political or commercial considerations, and/or a lack of prioritisation, investment, resourcing and contextual understanding, particularly in the Global South. Over recent years companies have come under increasing political, media and civil society scrutiny and pressure across all three areas, and are increasingly likely to be subject to regulatory oversight over the coming years. *Note: Labour Rights and concerns associated with companies’ outsourcing of human moderation are covered in the Annex.*

Democracy Features	**Rule of Law / Separation of Powers / Independent Judiciary**	**Transparency / Accountability in public admin.**	**Free & Fair Elections**	**Pluralistic political system**	**Free / plural media**
Impact / Why Important?	In most contexts private sector content moderation lacks state or regulatory oversight and therefore due process, which limits opportunities for genuine independent appeals and redress when rights may have been infringed. However, there are also significant concerns of too much state control over speech in both democratic and non-democratic contexts. Any regulatory oversight of content moderation must be independent of government and/or political interference to ensure platforms are not intentionally biased in favour of any particular political actors (i.e. current governments).	Private sector content moderation lacks transparency and accountability in terms of their policies and processes, and the extent to which outcomes are impacted by commercial, political or reputational considerations. While social media platforms are private, they conduct moderation activities that would usually be done via the legal process (e.g. determining whether content is illegal), and can impact rights (e.g. via censorship). Outside of formal legal requests, currently there is limited transparency around governments' interactions with companies, for example when pressuring them to moderate certain types of content (either illegal or legal). Transparency is vital to understand the states' influence on freedom of expression and other key rights.	Content moderation policy has come under heightened scrutiny during key political moments such as elections as companies have had to moderate political advertising, a range of online harms, and sometimes electoral candidates and parties themselves. There have been concerns of both suppression of content and/or insufficient moderation by companies during elections, and an overall lack of consistency and resourcing in many contexts. Again, any regulatory oversight of social media during elections must be independent from political interference.	Effective and consistent moderation policies and practices should enable broader political participation if they achieve the desired outcome of reducing online harms, and the chilling effect that they have on online speech, especially for marginalised or minority political voices. However, ineffective and inconsistent moderation policies and practices may limit political participation and plurality by forcing marginalised or minority voices out of online spaces, and reporting and flagging systems can be open to abuse and used to suppress certain political positions or groups.	Effective and consistent moderation policies and practices should help to protect the quality of the media ecosystem by limiting the reach and impact of dis/misinformation, the abuse or harassment of journalists etc. However, there is a danger of regulation in this area being used to suppress dissenting media through government pressure, for example in contexts such as Turkey or Hungary, highlighting the need for any regulation to be fully independent. The automation of content moderation (often partially as a result of government pressure) can also impact media or civil society reporting on sensitive issues such as war crimes or terrorism, with content being incorrectly censored.
Human Rights	**Non-discrimination, Minority protections, Incitement, Security**	**Freedom of Expression / Belief / Association / Political Participation**	**Privacy / Defamation**	**Freedom of information**	**Necessary conditions & limitations**

Impact / Why Important?	Effective content moderation policies and practices are required to limit the variety of online harms that affect minority or marginalised communities, including hate speech, dis/misinformation or abuse, but also discriminatory moderation of certain groups – e.g. when documenting offline violence (see below).	Achieving an effective balance between over and under-moderation requires an awareness of both freedom of expression concerns, especially if content is legal (but might be harmful), and the chilling effect on political participation and freedom of expression for groups impacted by online harms.	Effective moderation is required to combat online abuse that violates privacy rights (e.g. doxing, non-consensual sharing of images etc.), or dis/misinformation that may be defamatory. Moderation policies and practices should also safeguard the privacy of users that report or flag content or behaviour, to ensure their privacy is protected from those they are reporting. Privacy laws are also vital to protect users' data from unwarranted government or law enforcement intrusions, and companies should only share user data when required to do so through legitimate, formal legal processes (see Privacy section below).	Content moderation by private companies, especially when applied incorrectly, can interfere with or limit the flow of legal content, potentially infringing on users rights to freely access and share information.	Governments may only constrain rights such as freedom of speech and freedom to information to protect the rights of others, and under specific exemptions such national security or public health. Any government intervention that impacts these rights (via regulation of content moderation) must be legal, necessary and proportionate (see Foundational Principles), and conducted by an independent regulator to prevent political interference. Under existing intermediary liability regimes (see below), the same restrictions do not necessarily apply to private companies in the same way, raising questions over whether private companies with dominant market and societal positions should wield such power over fundamental rights without greater transparency, oversight and accountability.

Online Harms	**Disinformation & Misinformation**	**Hate Speech & Incitement**	**Online Abuse & Harassment**	**Online censorship**	**Invasions of Privacy**
Impact / Why Important?	Effective and consistent moderation is required to limit the impact of dis/misinformation, hate speech, and online abuse and harassment by preventing the amplification of harmful content, removing harmful accounts, and providing a safe online environment with clear and enforceable rules. In the most serious cases, under-resourced and ineffective content moderation has fuelled offline violence. Moderation therefore requires significant resources to be effective at the scale of online platforms, and sufficient local knowledge and language expertise to ensure moderation is fair and consistent across different contexts. A lack of resources and expertise in content moderation can also lead to incorrect moderation decisions, for example through an over-reliance on insufficiently accurate AI tools, which can have a negative impact on those trying to prevent			Effective and consistent moderation is also required to prevent over-moderation or censorship. When content moderation is conducted by private companies under current intermediary liability arrangements and without regulation, they lack due process, opportunities for redress, transparency and accountability. This can disproportionately impact those in the Global South where companies have invested far fewer resources in content moderation.	Effective moderation is required to combat online abuse that violates privacy rights, e.g. doxing, non-consensual sharing of images etc. The use of AI tools in content moderation can lead to potential invasions of privacy, for example through the 'general monitoring' of all content, or scanning user content pre-upload. Privacy laws are also vital to protect users' data from unwarranted government or law enforcement intrusions, and companies should only share user data when required to do so through legitimate, formal legal processes (see

	online or offline harms, e.g. documenting human rights abuses during a conflict.		Privacy section below).
What does good look like?	**Governments:** - Any regulation that impacts social media platforms' content moderation must be independent from government and any political interference, and meet key certain criteria to comply with human rights (see above). - Governments should focus on content moderation policies and systems and their overall impact on democracy, human rights, and online harms, rather than on specific cases.This could include requiring effective risk assessments and due diligence and resourcing from companies, combined with genuine opportunities for appeals and redress, overseen by a regulator with sufficient resources and expertise. - Any regulation should start with a focus on improving transparency to better understand the impacts of social media platforms' content and recommendation systems, including data access for regulators and third-party experts.	**Companies:** - Companies' content moderation policies should protect democracy and human rights, be clear and consistent, independent of political or commercial considerations, and informed by contextual understanding and consultation, particularly in the Global South, with marginalised or minority communities, and during crises or conflicts. - Internal company processes and practices for content moderation should be fully resourced across all contexts to manage the scale and speed of moderation required on their platforms in each market, and have effective prioritisation to prevent the most serious forms of online harm (e.g. incitement) that could lead to offline violence. - Company enforcement decisions and actions must be as consistent as possible to limit under and over-moderation and the resultant impacts on democracy and human rights, provide clear explanations of outcomes, and fair opportunities for appeals. - Companies should be as transparent as possible across all three areas, and iteratively update their policies and practices to improve content moderation outcomes. - Companies should address concerns around the labour rights of workers conducting content moderation on their behalf (e.g. pay and conditions, psychological impacts etc.), both internally but especially external where moderation is outsourced to other companies.	
Regulation *See 'Levers' section of framework for key considerations checklist for regulation*	**Key considerations:** - Should content moderation decisions be left to private companies to effectively police speech and determine the legality (or harmfulness) of content without formal legal processes or independent oversight? If so, then regulation should ensure there is clarity, consistency, fairness, and effective systems for appeals and redress, and that these systems are sufficiently resourced and transparent. If not, can state institutions and systems (law enforcement, justice systems) be designed and operated in a way that can deal with the scale and scope of online content? - Regulation of 'legal but harmful' content remains a highly contested area. Any regulation should treat illegal and 'legal but harmful' content differently, with clear but distinct approaches to ensure that legal content (even if harmful) is not de-facto criminalised to protect freedom of speech. Proposed approaches	**Examples** **Content-Focused:** **German NetzDG** Germany fines Facebook €2M for violating hate speech law (Politico) Germany's balancing act: Fighting online hate while protecting free speech (Politico) UN Human Rights Committee Criticizes Germany's NetzDG for Letting Social Media Platforms Police Online Speech (EFF) The Digital Berlin Wall: How Germany (accidentally) Created a Prototype for Global Online Censorship (Justitia, Sept 2020). **French Dis/Misinformation law: 'Against information manipulation'** French Parliament passes law against 'fake news' (Politico) French MPs criticise 'hasty and ineffective' fake news law (Guardian)	

	to dealing with ‘legal but harmful’ content typically focus on ensuring that private companies have clear policies (to either permit or not) these types of content, and that their moderation processes consistently enforce them to ensure predictable outcomes for users. To protect freedom of speech, companies could be encouraged to take actions that stop short of content removal (e.g. not amplifying legal but harmful content via recommendations or newsfeeds, adding fact-checks etc.).	**French Hate Speech law (‘Avia Law’)** France gives final green light to law cracking down on hate speech online (Politico) French Court Strikes Down Most of Online Hate Speech Law (NYT) France’s watered-down anti-hate speech law enters into force (Universal Rights Group) **Systemic-Approaches:** Moderating online content: fighting harm or silencing dissent? (UN OHCHR, July 2021) **EU Digital Services Act** Europe's Digital Services Act: On a Collision Course With Human Rights (EFF) The Digital Services Act could require big changes to digital platforms. Here are 4 things lawmakers need to know to protect people-powered spaces like Wikipedia (Wikimedia, Nov 2021) How can we apply human rights due diligence to content moderation? Focus on the EU Digital Services Act (CDT) Online Platforms' Moderation of Illegal Content Online: Law, Practices and Options for Reform (EP IMCO Committee, June 2020) **UK Online Safety Bill** ‘Legal to Say, Legal to Type’ Campaign & Consultation Submission (Index on Censorship) UK's Draft Online Safety Bill Raises Serious Concerns Around Freedom of Expression (EFF) Online Safety Bill: Five thoughts on its impact on journalism (LSE) How the OSB Lets Politicians Define Free Speech (Open Rights Group) UK OSB: Harmful (but legal) content, and what’s next (Herbert Smith Freehills)
Non or Self-regulation *See ‘Levers’ section of framework for*	**Key considerations:** - If regulation is not a feasible or desirable option, then pressuring companies to more effectively enforce their policies and meet their existing commitments (e.g. human and civil rights audits, risk assessments) in an equitable and consistent way may be the only option to improve content moderation outcomes. This should include investing sufficient resources across all contexts in which companies operate to ensure equitable outcomes that are	**Examples** **Int. Orgs:** Joint Declaration on Freedom of Expression and the Internet (OSCE, 2011) Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence (UNHR, 2013) Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation (Council of Europe, 2021)

key considerations checklist for non-regulation	proportionate to risks. - Companies should also be pressured to engage more constructively with civil society and communities affected by online harms to create an online environment that is equitable and inclusive of all communities and proactively prioritises protections for marginalised communities, rather than reacting to crises or high-profile incidents. - It is important to connect individual examples of content moderation failings to companies' broader content moderation systems and resourcing decisions to undermine justifications for these failings based on the scale of content online. For example, the FB Oversight Board has been criticised for only being able to look at individual cases, although the Board has included broader recommendations based on individual cases. - Self- or co-regulatory initiatives often represent baseline standards, especially when developed in partnership with big tech private sector companies, as they have been unwilling to commit to more ambitious or systemic approaches.	**Public-Private Partnerships:** WeProtect (CSEA-focused) **Global Internet Forum to Counter-Terrorism (GIFCT)** GIFCT Transparency Report Raises More Questions Than Answers (Brennan Center, Sept 2019) Human Rights NGOs in Coalition Letter to GIFCT (Various, July 2020) A Human Rights Assessment of the GIFCT (BSR, July 2021) **Christchurch Call** **Private Sector:** FB Oversight Board & Oversight Board demands more transparency from Facebook Does Facebook's Oversight Board Finally Solve the Problem of Online Speech? (CIGI) Facebook's Oversight Board makes an imperfect case for private governance (Brookings) Facebook and the Folly of Self-Regulation (Wired) Social Media Councils: One piece in the puzzle of content moderation (Article19) **Civil Society:** Santa Clara Principles On Transparency and Accountability in Content Moderation (Various NGOs)

Policy Area	Liability & Enforcement
Intro	Despite long-running debates and huge amounts of developing case law surrounding the liability provisions that govern online platforms, these arrangements have remained largely unchanged since the 1990s and early 2000s in many contexts. For example, Section 230 of the US Communications Decency Act (1996) and the EU's E-Commerce Directive (2000) have provided the legal basis for much of the growth we have seen in online social media platforms over the past two decades, by exempting such platforms from 'intermediary liability' for content created and uploaded by third-parties (also sometimes referred to as 'safe harbour' or 'mere conduit' provisions in different contexts). As such, online platforms have not faced the same levels of liability as traditional publishers or media platforms for content they 'host' but do not create or 'publish' directly themselves. These laws also protect the rights of these platforms to moderate the content they host as they see fit (including legal content), with some exceptions for certain types of illegal content once they become aware of its presence on their platforms (e.g. CSEA/CSAM, copyrighted material). The rationale for these approaches to liability was that freedom of speech and a thriving internet economy could only be guaranteed if online intermediaries were not held liable for the user-generated content hosted on their platforms. However, these liability arrangements currently provide little or no legal incentive for social media platforms and companies to better prevent or mitigate harms to democracy and human rights facilitated by their platforms, beyond reputational or brand considerations (threatening their advertising revenues) and political and media pressure (threatening additional liability via new legislation or regulation). In fact, in some instances the prohibitions on 'general monitoring' (i.e. indiscriminate or mass assessments of all content uploaded to a platform), and the imposition of liability only once a company is aware of certain content, have created a disincentive for companies to proactively monitor or moderate some forms of online harm. Over the past decade various self- or co-regulatory initiatives have emerged in an attempt to encourage online platforms to tackle both illegal and 'legal harms', with varying but typically limited success. Despite some marginal improvements through informal, voluntary or industry-led approaches, over recent years the underlying liability arrangements have come under increasing pressure as the range, scope and persistence of both illegal, and 'legal but harmful' content and activities on online platforms has been exposed, resulting in a raft of proposed changes to liability arrangements across different contexts. In general, this emerging trend towards new online regulation can be divided into two broad categories: ● *Content-based approaches*, such as Germany's NetzDG laws, often targeting a specific online harm such as hate speech or electoral disinformation, and focusing on the effective and timely removal of that content where appropriate. ● *Systemic approaches*, such as the EU DSA or UK Online Safety Bill, whereby online platforms must demonstrate that their policies, processes and systems are designed and implemented with respect to the risk of negative outcomes that could occur, across a range of possible harms. The introduction of new online regulations has opened up a series of debates around the requirements of the regulatory bodies tasked with overseeing and implementing these new liability arrangements, including around the levels of resources, access, investigative powers and sanctions required to play this role effectively (e.g. fines, reduced liability protections, geo-blocking, individual liability for senior company leadership and/or management etc.). New regulations also raise a series of potential challenges, from resourcing and expertise constraints to jurisdictional issues, the explicability of automated systems, and the difficulty in keeping pace with the evolution of both online platforms and harms (i.e. the impact of any enforcement is retrospective, rather than preventative).

<table>
<tr><td></td><td colspan="9">Finally, as outlined in further detail in the Regulatory & Non-Regulatory Levers section, it is important to note that additional liability or enforcement powers could be potentially dangerous or counter-productive in non-democratic contexts, and there are also dangers associated with setting precedents in democratic contexts that could then be abused elsewhere (e.g. geo-blocking leading to online censorship, senior leadership or management liability for speech, abuse of access to company data for political persecution etc.).</td></tr>
<tr><td><b>Democracy Features</b></td><td colspan="2"><b>Rule of Law / Separation of Powers / Independent Judiciary</b></td><td colspan="2"><b>Transparency / Accountability in public admin.</b></td><td colspan="2"><b>Free & Fair Elections</b></td><td colspan="2"><b>Pluralistic political system</b></td><td><b>Free / plural media</b></td></tr>
<tr><td><b>Impact / Why Important?</b></td><td colspan="2">Current liability frameworks leave an accountability gap, relying on private sector enforcement by platforms with limited government or state oversight or scope to appeal, meaning that citizens' relationships with companies are not fully governed by rule of law processes. However, a lack of effective enforcement of illegal content (e.g. hate speech) or activity (e.g. abuse and harassment) online also creates a rule of law deficit, as crimes go unpunished and impacted victims do not receive justice.
This has begun to change, first with content-focused regulations that formalised reporting and appeals processes (e.g. NetzDG), and more recently with more systemic approaches. However, both types of approach have been criticised for enshrining the primary role of private companies in content moderation, including determinations on whether content is illegal or protected speech, rather than courts following the rule of law.
There are also concerns about digital regulations being abused by governments to suppress free speech, especially in contexts where there is limited separation of powers between the government and the legal system and/or regulator.</td><td colspan="2">Both content-based and systemic approaches to digital regulation that alter current liability regimes typically included new or additional transparency and oversight mechanisms for companies (see Transparency section above). However, it is vital that they also include sufficient transparency and accountability provisions for the development and implementation of the regulations, for example from government departments or state regulators.</td><td colspan="5">Updates to liability regimes should lead to more consistent and effective enforcement responses from companies, and thereby reduce the impacts of online harms across different features of democracy, from elections to political participation and media freedoms. However, content-focused or systemic approaches that only cover illegal content would not have the same impact on harms that typically do not break the law (e.g. disinformation).
In the context of elections, current liability regimes have meant that company responses to online harms <i>during</i> elections have often been too slow or reactive to provide effective protection during the election period itself, with action and/or sanctions typically applied retrospectively (e.g. takedown of disinformation networks, abuse of political ads and/or campaign finance violations). Systemic regulation that encourages more proactive approaches should help to speed up action during the election cycle.
Additional regulation should also help to foster broader political participation online, and prevent harms targeting the media online. However, there are also concerns over the potential impacts on freedom of expression and the media if increased liability is used as a tool by governments to target political opposition, or if it creates incentives for companies to over-moderate. For example, in the EU there have been debates over whether media content should be exempted or not.</td></tr>
<tr><td><b>Human Rights</b></td><td><b>Non-discrimination, Minority protections, Incitement, Security</b></td><td colspan="2"><b>Freedom of Expression / Belief / Association / Political Participation</b></td><td colspan="2"><b>Privacy / Defamation</b></td><td colspan="2"><b>Freedom of information</b></td><td colspan="2"><b>Necessary conditions & limitations</b></td></tr>
</table>

Impact / Why Important?	Increasing the legal responsibility of companies to effectively moderate content and behaviour on their platforms should create a stronger legal incentive for companies to address security related human rights violations and better protect minorities from discriminatory or inciting content online.	However, increasing the legal responsibility of companies to moderate content and behaviour through regulation also has the potential to create an incentive for companies to over-moderate to avoid potential liability, thereby negatively impact on freedoms of expression, especially for ‘legal but harmful’ content. Increasing platform liability also has the potential to cause jurisdictional issues. For example, if a company does not comply with regulations in one country, and is based in another, then regulators would have few options to force compliance. In serious cases of non-compliance, this could result in entire platforms being geo-blocked by regulators in that country, raising questions of proportionality even if the platform is predominantly toxic or harmful (but not illegal). From a rights perspective, it is also vital that any regulation does not include overly broad powers that could allow for political interference in regulatory scope or decisions, e.g. the powers for government ministers in the UK’s draft OSB.	Additional liability for companies could have a variety of impacts on privacy rights: For example, greater liability could create incentives for companies to require confirmation from users of their identity to crack down on anonymous abuse and harassment, thereby protecting the privacy of those receiving such abuse, but undermining privacy safeguards overall. Alternatively, greater liability could also lead companies to introduce further encryption so that they do not have access to content, and therefore cannot be held responsible for moderating it, even though this would also strengthen user privacy.	Increasing the involvement of the state in regulating communications has the potential to impact freedom of information through political interference, or the use of liability laws to control the public sphere.	Liability regimes that include provisions on ‘legal but harmful’ content may not meet the criteria required under human rights standards for laws that impact rights such as freedom of expression (e.g. legitimacy, proportionality etc.). Even if the content moderation is conducted by companies themselves, liability regimes that demand this may still represent states overreach.

Online Harms	**Disinformation & Misinformation**	**Hate Speech & Incitement**	**Online Abuse & Harassment**	**Online Censorship**	**Invasions of Privacy**
Impact / Why Important?	Overall stronger regulation that increases the potential liability and enforcement faced by companies should result in better prevention and mitigation of these types of online harm. However, the extent of this will depend on whether this includes just illegal content (in which case dis/misinformation would not be covered), or also ‘legal but harmful’ content, and whether this would apply for all services and types of users (e.g. adults vs. children). New regulatory regimes that amend current liability arrangements are also likely to include new powers to require additional transparency and data access from companies (whether purely to regulators, or also third-parties such as academics and civil society,			Conversely, increasing the level of liability faced by companies also has the potential to create incentives to over-moderate, especially if ‘legal but harmful’ content is included. Additional liability also increases the potential for state or political interference, with more powers to compel companies to moderate	See Privacy above.

	and the public). This additional level of oversight should also over time enable a more sophisticated and complete understanding of the prevalence and impacts of these types of harm.	content in a certain way (see Freedom of Expression above).	

What does good look like?	**Governments:** Any regulation that impacts social media platforms' liability must be independent from government and any political interference, and meet key certain criteria to comply with human rights (see above). Governments should focus on companies' policies and systems and their overall impact on democracy, human rights, and online harms, rather than on specific content moderation cases.This could include requiring effective risk assessments and due diligence and resourcing from companies and an emphasis on safety by design, combined with genuine opportunities for appeals and redress, overseen by a regulator with sufficient resources and expertise.	**Companies:** Not applicable as companies will not voluntarily take on further legal liability, but without regulation pressure can be applied on companies to agree to self or co-regulatory standards, and cross-industry cooperation to share resources and best practices where possible. If regulation is in place then companies should engage and cooperate with regulators in a constructive and transparent manner, and again work across industry where possible.
Regulation *See 'Levers' section of framework for key consideration s checklist for regulation*	**Key considerations:** - Liability is a complex and fundamental area of internet regulation that often provides the foundations for additional provisions in other policy areas. Overall, approaches to liability that focus on companies systems and processes appear to be gaining more traction than narrow content-focused approaches, particularly as they have the potential to keep pace with the scale and evolution of online platforms, and aim to create sufficient incentives for companies to balance online safety and freedoms. - As outlined in the Levers section of the framework, context is crucial when determining the potential impact of a given regulatory approach on democracy and human rights, and increasing the liability of companies in non-democratic contexts may present further risks to human rights. As outlined above, any regulation in this area must be independent and free from government or political interference. - Changes to liability must also be cognisant of competition and innovation concerns, and ensure that they are proportionate to the scale and impact of harms on different platforms and types of online services otherwise there is a risk of further entrenching the largest companies with already dominant market positions that can best afford to comply with new regulatory obligations..	**Examples** **Content-focused:** The push for content moderation legislation around the world (Brookings, Sept 2020) **German NetzDG** NetzDG Legal Analysis (Article19, 2017) The Impact of the German NetzdG law (CEPS, 2018) Germany is amending its online speech act NetzDG... but not only that (Internet Policy Review, April 2020) **Systemic:** **EU Digital Services Act** The Digital Services Act: What are the key provisions, and does it strike the right balance? (Lexology, May 2021) The Digital Services Act: From Intermediary Liability to Platform Regulation (Miriam Buiten, June 2021) "Too Big to Care" or "Too Big to Share": The Digital Services Act and the Consequences of Reforming Intermediary Liability Rules (ECIPE, April 2021 - *Note: company funded org - reflects industry positions*) The proposed DSA – Part 3: The liability exemptions and the 'notice and action'

		mechanism (Field Fisher, April 2021) **UK Online Safety Bill** The Draft Online Safety Bill: Carnegie UK Trust initial analysis (Carnegie, June 2021) Joint Scrutiny Committee Evidence Submissions (Various, 2021) The draft Online Safety Bill: systemic or content-focused? & Online Harms Compendium (Cyberleagle) Online Abuse: Why Management Liability Isn't The Answer (Open Rights Group) & The UK's Online Safety Bill is a "hostage-taking" law. That should terrify you (Heather Burns) National & International Models for Online Regulation; The EU Digital Services Act & the UK Online Safety Bill; Future Considerations for Online Regulation (ISD, 2020) How Other Countries Have Dealt With Intermediary Liability (ITIF, 2021)
Non or Self-regulation	**Key considerations:** NA: Any changes to liability and enforcement regimes require amendments to existing legislation or regulation, or new legislation, although there are some sets of principles that have been established by NGOs/civil society.	**Examples** Manila Principles on Intermediary Liability (EFF et al, 2015) Nine Principles for Future EU Policymaking on Intermediary Liability (CDT) Design Principles for Intermediary Liability Laws (Transatlantic Working Group) Intermediary Liability & Content Regulation (GNI)

Policy Area	Privacy & Data Protection				
Intro	Data protection is a broad area of policy that reaches beyond ad-tech companies, with regulations such as the EU's GDPR (2018) covering any company, organisation or state institution that collects, processes or shares individuals' personal data to prevent infringements of rights to privacy (i.e. freedom from illegal or unauthorized intrusion). Considered among the most stringent, GDPR has spurred an increase in new laws resulting in a patchwork of regulation of varying strengths internationally, and in some contexts there remains little if any regulation in place. Data protection regulations are typically constructed around informed consent, allowing individuals to decide how their data is used, and by who. However, consent is typically a condition of access to many online services or content, and (legal) consent is often secured through processes marked by stark information and power differentials (e.g. lengthy and highly legalistic and inaccessible Terms of Service, 'dark patterns' in UX design) that call into question the possibility of *meaningful* consent. Similarly, the dominant market position of many of the largest ad-tech companies, and the increasing centrality of their services in communications (e.g. WhatsApp), business (e.g. GMail, Google Suite etc.) and finance (e.g. mobile FB payments), mean that in many contexts citizens have little choice but to consent to their terms. Data protection regulation is therefore intrinsically connected to ad-tech companies' business models that rely on the collection and processing of user data to target advertising at specific subsets of online users. Regulation often also includes transparency provisions, for example requiring those collecting or processing data to share any data they hold on an individual under 'Subject Access Requests'. As the internet and mass data collection have become pervasive over the past several decades, there have been numerous examples of this abundance of data being exploited by governments to conduct indiscriminate mass surveillance by both democracies and authoritarian states, ostensibly to protect national security (albeit with varying levels of legitimacy, oversight or accountability). Recent years have also seen a blurring in the line between state and non-state infringements on privacy online, such as state hacking using tools developed by the private sector. As with data protection regulations, in many contexts there has been a push to update the surveillance laws that govern access to data by governments and law enforcement agencies. Under certain limited circumstances governments or law enforcement agencies may intercept personal data, surveil individuals online or compel information from companies. Any state infringements on privacy must be legal, necessary, proportionate, and have 'adequate safeguards' to not violate privacy rights. Privacy has also been central to debates around online anonymity and encryption, and the distinctions between public and private spaces online, with some governments and law enforcement agencies pushing for greater access to private communications and/or weakening of end-to-end encryption (E2EE) to combat illegal activity such as terrorism, CSEA or hate speech online by making it easier to indentify perpatrators. However, as many privacy activists have argued, online safety and privacy are not mutually exclusive, and in many contexts strong privacy protections and tools to maintain online anonymity are essential to the security of political dissidents, minority or marginalised groups. Looking ahead, data protection regulations will need to keep pace with technological change, with both privacy (e.g. decentralised social networks, enhanced encryption etc.) and surveillance tech (e.g. via a growing surveillance tech sector) continuing to evolve.				
Democracy Features	Rule of Law / Separation of Powers / Independent Judiciary	Transparency / Accountability in public admin.	Free & Fair Elections	Pluralistic political system	Free / plural media

Impact / Why Important?	Privacy protections are vital, especially in contexts where these features of democracy are not present to protect individuals from the state, and in the most serious cases privacy violations can lead to offline violence by state or non-state actors. Rule of law and an independent legal system is vital to ensure the policing of illegal content and activity online respects human rights, protecting individuals (and companies) from state overreach or political interference in requests for access to data. Many companies operate in contexts where these features are not present, and have been accused of helping authoritarian governments censor online speech.	Transparency and accountability for government or law enforcement data requests to companies vital to understand scale, proportionality and legality of state surveillance and censorship. Transparency and accountability vital for oversight of how ad-tech companies collect and use individuals' data, and for investigations into any privacy or data breaches. Technology designed to protect users' privacy can also be used to undermine transparency and accountability by governments or state actors, for example through the use of encrypted and/or disappearing messages.	The use of data by political campaigns has led to some of the most high profile examples of breaches of user privacy involving ad-tech companies (see Advertising). Effective data protection is therefore vital to ensure fair campaigning during elections, but also to protect against the surveillance of political opposition, hack and leak tactics etc.	Invasive data collection can enable violations of privacy through surveillance or hacking (either from domestic or foreign governments, or non-state actors), which can have a chilling effect on political dissent, and/or lead to a variety of online harms (dis/misinformation, abuse, harassment, offline violence) targeting political opposition, activists, civil society etc.	Infringements of privacy, such as the hacking of journalists, has significant negative impacts on media freedoms and a chilling effect on reporting. Privacy through anonymity is also essential for protecting journalistic sources (e.g. whistle-blowers) that are vital to a free press. Citizens' data however should also be protected from media intrusion, unless there are strong public interest grounds to report on hacked, leaked or stolen data.
Human Rights	**Non-discrimination, Minority protections, Incitement, Security**	**Freedom of Expression / Belief / Association / Political Participation**		**Freedom of information**	**Necessary conditions & limitations**
Impact / Why Important?	Data protection regulations and privacy/anonymity tools are vital for individuals' security in many contexts, whether democratic or not, especially for persecuted minorities, political opposition, dissidents and activists. Data protection is vital to prevent infringements of privacy leading to discrimination, incitement, or offline violence (e.g. 'outing' of LGBTQ people, doxing etc.), and limit the collection of personal data that increases these	Surveillance and other infringements of privacy can be a threat to individuals' security, and by extension all of these fundamental rights. Anonymity can therefore be vital for freedom of expression and political participation in many contexts, especially for minority or marginalised communities to avoid various forms of online harm e.g. hate, abuse etc. However, these same protections can be abused to protect perpetrators of online harms that impact the exercise of these rights (e.g. hate speech and incitement,		Privacy rights must be balanced with rights to freedom of information, for example where there is a legitimate public interest in releasing private data or information. This tension is most prominent with 'right to be forgotten' or 'right to erasure'	Any limits on privacy need to be limited, necessary and proportionate, and follow rule of law and due process to be legitimate under human rights law. Open to abuse to justify (mass) surveillance on national security grounds in both democracies and

	vulnerabilities. Data protection is vital to restrict ad-tech companies' mass data collection and limit associated harms – see Advertising.	dis/misinformation targeting certain groups, abuse and harassment).	provisions that are included in privacy laws such as GDPR, which address issues such as how long information on spent criminal convictions should remain public.	authoritarian contexts, but more typical in the latter to suppress political opposition, marginalised or minority communities (e.g. Uighurs in Xinjiang).
Online Harms	**Disinformation & Misinformation**	**Hate Speech & Incitement**	**Online Abuse & Harassment**	**Online Censorship**
Impact / Why Important?	Infringements of privacy often play a key role in various online harms, from hacked or manipulated information being used as the basis for disinformation, to the use of private information to fuel online abuse (e.g. doxing, non-consensual sharing of images etc.). Overall strong privacy protections are vital in constraining the worst outcomes of ad-tech business models that rely on mass data collection, as these practices heighten vulnerabilities to a range of harms, particularly for minority or marginalised communities.			Privacy protections are also important in the context of online censorship, as governments and/or law enforcement may request individuals' personal data from companies on those whose content is deemed illegal and removed. While this may be desirable in certain instances (e.g. identifying those creating, hosting or sharing CSEA content) and contexts (i.e. where rule of law processes are followed, courts are independent from political interference etc.), such data requests can also be used to identify dissidents or political opponents, or enforce laws that may not be compatible with rights to freedom of speech (e.g. blasphemy or 'insult' laws).
What does good look like?	**Governments:** - Governments should ensure legal frameworks are in place that comprehensively safeguard privacy and personal information, based on informed consent, and do not undermine anonymity online. - Individuals should have control over their data, including the right to access it, amend or delete it, and easily transfer it between different companies or services when needed. Governments should also ensure that citizens are aware of their privacy and data rights, for example through public education or awareness. - Regulations should require companies to minimise the collection of user data as far as possible, and clearly disclose in plain language how they use it. Companies should also be required to disclose the details of any third parties that have access to the data, and for what purposes. Companies should also be required to notify customers in a timely fashion if their data is compromised.			**Companies:** - Companies should provide users with clear and accessible explanations of what data is collected, how it's used, and who it's shared with. - Companies should adhere to privacy by default (or design) and ethical design principles, especially for services used by children, and not employ deceptive or manipulative design practices to impede user choices. - Companies should adhere to data minimisation principles, and limit data collection and retention to what is required for a specific purpose. In the context of ad-tech platforms, this principle would limit the types of data that companies use to target advertising at users (and resultant harms, e.g. discrimination). - Companies should ensure that any government data requests they receive are legitimate and adhere to international human rights standards, and provide thorough transparency on these requests. Companies should not comply with or

	- Data protection regulations should limit the types of sensitive personal data that can be collected in the first place (e.g. on protected characteristics) to protect from discriminatory practices, as well as limit how data can be combined, processed or shared to infer additional data points (e.g. infering protected characteristics from other data points, e.g. interests, location data etc.). - Data protection regulations should include strong transparency and accountability provisions that cover government or state data handling and requests, the private sectors' collection and use of data, as well as transparency for the regulator or body charged with enforcing the regulations. - Effective data protection regulation also requires sufficient investigative and enforcement powers (and resources) to create the right incentives for governments, state bodies and private sector companies to comply, alongside appeals and redress mechanisms that respect the rule of law.	fight illegitimate requests, and carefully consider the data protection and privacy implications of operating in countries without adequate democratic and human rights safeguards.
Regulation *See 'Levers' section of framework for key considerations checklist for regulation*	**Key considerations:** - With the raft of new privacy and data protection regulations introduced worldwide over the past few years, governments should continue to work towards strong international standards to promote best practices in regulation. This should help to ensure that there are not widely different approaches across national or regional jurisdictions, undermining the open nature of the internet. - Privacy and data protection regulations often apply across different sectors and industries, and also overlap with other areas of regulation (e.g. elections, advertising, consumer protection etc.) so require effective coordination between different regulators and state bodies. Resourcing and enforcement of data protection regulations must be strong enough to provide an effective incentive for companies to comply and if necessary, change their business practices. - Privacy and data protection legislation and regulation can be abused as a tool of repression (particularly although not exclusively in non-democratic contexts), for example by providing overly broad powers for the government or law enforcement to request citizens' data from companies, requirements for 'data localisation', or 'backdoors' into encryption or company servers.	**Examples** **EU GDPR** European Commission Two Year Evaluation Report Ireland's Facebook decision triggers argument over limits of GDPR (Politico, Oct 2021) **California Consumer Privacy Act** California's groundbreaking privacy law takes effect in January. What does it do? (Guardian, Dec 2019) The California Consumer Privacy Act ("CCPA") – 2020 Year in Review (National Law Review) Data Protection and Privacy Legislation Worldwide (UN CTAD Database) **German NetzDG** Germany tightens online hate speech rules to make platforms send reports straight to the feds (TechCrunch, June 2020) The future of data protection: what we expect in 2021 (AccessNow)

Non or Self-regulation *See 'Levers' section of framework for key considerations checklist for non-regulation*	**Key considerations:** - Tech companies are unlikely to voluntarily change their approaches to data protection and privacy given their intrinsic connection to the ad-tech business model, where user data is essentially the product sold via ad targeting. - However, other tech companies (e.g. Apple, DuckDuckGo) use privacy as a selling point of their products, which can create competition and pressure within the tech industry. - There are some existing self-regulatory initiatives, such as GNI, and privacy standards set by international organisations (see Examples).	**Examples** GNI Principles on Freedom of Expression and Privacy OHCHR and privacy in the digital age & International Standards List (UN) OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (2013)