FAIR VOTE

Briefing: Joint Committee on the Online Safety Bill's Report

Implications of the report content on democracy and elections
14.12.21

Summary:

Earlier in 2021 the government of the United Kingdom published a draft of its proposed Online Safety Bill. Because of the complexity of the bill, a joint committee was established to perform pre-legislative scrutiny with the aim of improving the bill before it is tabled. On 14 December, 2021, The Committee's report was published. It importantly acknowledges the systemic nature of online harms, particularly as they pertain to democracy. They note that "Algorithms, invisible to the public, decide what we see, hear and experience" and recommend a systems approach over a content approach to regulating social media.

As it pertains to democratic harms, this report offers recommendations for robust Ofcom enforcement policies that will reduce disinformation and misinformation, reckon with anonymous accounts, ensure much-needed transparency, and curb the virality of content that is likely to cause significant harm at scale through mitigation.

Fair Vote UK is supportive of many key aspects of the report including:

- Recommending a systems-based approach to regulation;
- Retaining robust protections for freedom of expression;
- Acknowledging the harm caused by disinformation and misinformation on democratic processes and social cohesion including that which "is likely to undermine the integrity and probity of electoral systems";
- Putting safety by design at the core of regulation, with a duty on platforms to mitigate against foreseeable risks of harm arising form design, including "friction-increasing measures to slow down sharing, required moderation of groups over certain size, limit the number of "one click" shares possible, special arrangements for "periods of heightened risk" (ie elections)"
- Bringing paid-for ads into scope and recommending limitation of data use without explicit consent for targeting ads;
- Giving users the right to verify their account and the right to filter interactions with anonymous accounts, ensuring vulnerable users can still engage anonymously while empowering users to protect against bullying and harassment driven by anonymous accounts;
- Endorsing the Law Commission's recommendations for new criminal offences in its reports, Modernising Communications Offences and Hate Crime Laws to protect marginalised groups;

- Constraining the Secretary of State's unilateral powers, ensuring no government retains absolute power to determine the rules that govern the digital world;
- Requiring Ofcom to issue a mandatory code of practice to service providers on how they should comply with the duty of mitigating against content that poses societal harm at scale including prescriptive recommendations about context-competent human moderation and "dedicated teams for election periods and involve relevant bodies–with planned circuit breakers";
- Replacing existing protections around journalistic content and content of democratic importance with "a single statutory requirement to have proportionate systems and process to protect 'content where there are reasonable grounds to believe it will be in the public interest'";
- Rethinking the news media exemption to ensure bad actors - like those funded by foreign governments and those claiming to "citizen journalists" can be excluded from the concept of news publisher;
- Introducing binding minimum standards for the accuracy and completeness of risk assessments;
- Requiring robust transparency of platforms, including publicly accessible reports and "a statutory responsibility to commission annual, independent third-party audits of the effects of their algorithms, and of their risk assessments and transparency reports";
- Introducing board-level liability for failure to comply with their regulatory obligations when there is evidence of repeated and systemic failings.

While we broadly support the objectives of this bill with regard to child safety, fraud and other areas of online safety, this briefing is focussed primarily on the democracy and elections related areas of the report. They are outlined below in numbered sections that match those of the bill itself for easy reference. We have directly quoted or paraphrased the bill's language without losing meaning or intent. Items have been bolded by us for emphasis, with each bill section followed by an analysis from Fair Vote UK.

FAIR VOTE

Full Report Analysis:

Section Two: Objectives of the Online Safety Bill

Quotes and Paraphrases from the Report:

Harms affecting adults:

- Racist Abuse:
    - In many cases, the harm that these individuals face is direct abuse exacerbated or amplified by system design.
- Abuse against LGBTQ+ people:
    - Stonewall: 1 in 10 LGBTQ+ people experience online abuse directed specifically at them
    - The LGBT Foundation told us that LGBTQ+ people are also at risk of being harmed by the actions of platforms themselves, with LGBTQ+ content being erroneously blocked or removed at greater rates than other types of content.
- Misogyny:
    - Women are disproportionately affected by online abuse and harassment
- Religious Hate:
    - Antisemitism and islamophobia
        - Reset told us that, currently, "widely debunked far-right conspiracy theories about Islam run rife on social media sites/blogs", ranging from "claims of 'No Go Zones' in Western nations which are run by Sharia Law and bar non-Muslims and police" to claims about "a plot by Islamic nations to take over Europe to create 'Eurabia'
- Abuse against the disabled
- Freedom of Speech
        - Compassion in Politics described the "current climate of hostility, toxicity, and abuse online" and told us that this "prevents many people from joining social media sites". Their polling found that this can infringe on individuals' freedom of expression, with "1 in 4 … scared of voicing an opinion online because they expect to receive abuse if they do so"
        - DGM Media: "We believe it is incompatible with freedom of expression and media plurality for legitimate, responsible news content to be subject to blocking and take-down by a commercial organisation which is open to business pressures such as advertising boycotts, operates without due process, and has no authority to make judgments about the value of journalism."

Societal Harms:

- **Disinformation relating to democratic processes can affect social cohesion with societal divides having been exploited by malicious foreign actors to undermine democratic processes in the US and the UK.**

Factors Contributing to Societal Harms:

- "Quarterly reports from Alphabet Inc. and Twitter showed 92 per cent and 88 per cent of their respective profits were from advertising revenue"
- "We heard evidence from a range of sources that content that creates a risk of harm or factually inaccurate content is many times more engaging than innocuous or accurate content."
- Anti-defamation league: "When a user interacts with a piece of content, algorithmic systems recognise signals, like popularity, and then amplify that content. If content is forwarded, commented on, or replied to, social media algorithms almost immediately show such content to more users, prompting increased user engagement, and thus increasing advertising revenue."
- "Some people are also served disproportionately high amounts of content that creates a risk of harm by algorithms due to their personal characteristics, as inferred by the platform's algorithms."
- Prevalence Paradox: "some abusive posts, which make up a minority of content, are seen by a vastly disproportionately number of people."
- Black Box: One of the challenges of establishing exactly why content and activity that is abusive, false or creates a risk of harm is so overexposed is that the systems underlying platforms are like a "black box"

The Draft Bill and an Overarching Duty of Care:

- Concerns with the current draft:
  - Complexity
  - Lack of clarity around "journalistic content" and "content of democratic importance"
  - Lack of clarity around "person of ordinary sensibilities"
  - Lack of clarity around "priority content"
  - Too much power left to service providers
  - Transparency requirements not strong enough
  - Secretary of State powers undermine Ofcom's independence
- **<u>Recommendations:</u>**
  - **Restructuring of the bill to set out core objectives clearly at the beginning, and have everything else flow from those objectives. These will be based on the harms identified above.**

Fair Vote UK's Analysis – The Bill's objectives

- The report offers a comprehensive overview of harms to adults that, crucially, encapsulate how abusive and harmful content that can cause harm at scale is amplified and disseminated.
- There is a welcome focus on systems, design and algorithmic content recommendations that are core to the business models of social media platforms. The Committee importantly acknowledges that it is not in the best interest of digital platforms to regulate false or hateful content that generates more clicks and therefore more ad revenue. The committee effectively challenges the now disproved narrative that rates of abusive content being low by looking at *who actually sees it,* and acknowledges the startling lack of transparency around how algorithms actually function.
    - Their assessment of the current bill's problems and recommendations around restructuring the bill are essential if the legislation is to meaningfully deal with algorithmic and systemic content dissemination that undermines democracy.

<u>Section Three: Societal Harm and the Role of Platform Design</u>

Quotes and Paraphrases from the Report:

Content and Activity:

- The current draft is too focused on content and not enough on broader system design and activity.
- Service providers can, and should, be held accountable for carelessly hosting content that creates a risk of harm.
- One of the changes that the Government made in the draft Bill, compared to the White Paper, was to replace references to "content and activity" with references solely to "content". This has reinforced the sense among many of our witnesses that the draft Bill is concerned solely with content moderation.
- **<u>We recommend</u> that references to harmful "content" in the Bill should be amended to "regulated content and activity". This would better reflect the range of online risks people face and cover new forms of interaction that may emerge as technology advances. It also better reflects the fact that online safety is not just about moderating content. It is also about the design of platforms and the ways people interact with content and features on services and with one another online.**

Algorithmic Design:

- Platform design is central to what people see and experience on social media. Platforms do not neutrally present content. For most user-to-user platforms, algorithms are used to curate a unique personalised environment for each user.
- Designing curated environments for individual people can give them content that they are interested in and want to engage with, enhancing their experience on the platform. The commercial imperative behind this is to hold people's attention and maximise engagement. However, the choice to design platforms for engagement can be problematic:
- **"Engagement is maximised by (1) strong emotion, (2) rabbit holes that lead to a warren of conspiracy, (3) misinformation that gets engagement from detractors and supporters, and (4) … algorithmic reinforcement of prior beliefs."**
- Algorithms designed to maximise engagement can directly result in the amplification of content that creates a risk of harm
- Ms Haugen explained how recommendation systems can be designed to continuously serve content to people: "Instead of you choosing what you want to engage with, [YouTube] Autoplay chooses for you, and it keeps you in … a flow, where it just keeps you going."
- Frictionless activity:
    - Platforms are often designed to minimise friction for users, maximising their ability to interact with one another and diversify their communications through multiple different services with minimal effort. Autoplay, discussed above, is an example of a friction-reducing design feature
        - safety measures frequently come into conflict with the 'maximise engagement, minimise friction' incentives of the surveillance advertising business model."

Mitigation: Safety by Design

- DCMS described safety as: "the process of designing an online platform to reduce the risk of harm to those who use it … It considers user safety throughout the development of a service, rather than in response to harms that have occurred."
- Haugen noted the need to limit the amount of sharing possible
- Circuit breakers: "when content reaches a particular threshold of velocity or virality" you could send it to "the relevant teams within the platform so that they can assess what is happening"
- **<u>We recommend</u> that the Bill includes a specific responsibility on service providers to have in place systems and processes to identify reasonably foreseeable risks of harm arising from the design of their platforms and take proportionate steps to mitigate those risks of harm.** The Bill should set out a non-exhaustive list of design features and risks associated with them to provide clarity to service providers and the regulator which could be amended by Parliament in response to the development of new technologies.

Ofcom should be required to produce a mandatory Safety by Design Code of Practice, setting out the steps providers will need to take to properly consider and mitigate these risks. **Risks and mitigations include:**

- Algorithmic "rabbit holes"
    - Mitigations: transparency requirements, user control over algorithmic priorities, measures to introduce content diversity, allow people to deactivate recommendations
- Auto-play
    - Mitigation: limit auto-play and auto-recommendation
- Frictionless cross platform activity
    - Mitigation: warnings before following link
- Data collection and microtargeting
    - Mitigation: minimum requirements for transparency and placement and content of targeted adverts
- **Virality and frictionless sharing at scale**
    - **Mitigation: friction-increasing measures to slow down sharing, required moderation of groups over certain size, limit the number of "one click" shares possible, special arrangements for "periods of heightened risk" (ie elections)**

Anonymity:

- A common design feature across many user-to-user services is allowing anonymous and pseudonymous accounts, where users are either publicly unidentifiable or partly identifiable
- Some raised the possibility that verification did not have to be a mandatory process. They suggested numerous system design features that could address the risks posed by anonymous accounts. Clean Up the Internet argued that all users should have the option to verify their account and the option to control the level of interaction they have with unverified accounts on a sliding scale.
- Anonymous abuse online is a serious area of concern that the Bill needs to do more to address. The core safety objectives apply to anonymous accounts as much as identifiable ones. At the same time, anonymity and pseudonymity are crucial to online safety for marginalised groups, for whistleblowers, and for victims of domestic abuse and other forms of offline violence.
- **We recommend** that platforms that allow anonymous and pseudonymous accounts should be required to include the resulting risks as a specific category in the risk assessment on safety by design. In particular, we would expect them to cover, where appropriate: the risk of regulated activity taking place on their platform without law enforcement being able to tie it to a perpetrator, the risk of 'disposable' accounts being created for the purpose of undertaking illegal or harmful activity, and the risk of increased online abuse due to the disinhibition effect.

- **<u>We recommend</u> that Ofcom be required to include proportionate steps to mitigate these risks as part of the mandatory Code of Practice** required to support the safety by design requirement we recommended in paragraph 82. It would be for them to decide what steps would be suitable for each of the risk profiles for online services. **Options they could consider might include (but would not be limited to):**
    - **Design measures to identify large quantities of identical content coming from anonymous accounts**
    - **Governance patterns that assure human moderation of such situations**
    - **A requirement for large platforms to offer the choice of verified or unverified status and user options on how they interact with either category**
    - **Measures to prevent banned individuals from creating new accounts**
    - **Limit speed at which new accounts can be created and engage**
- **We recommend that the Code of Practice also sets out clear minimum standards to ensure identification processes used for verification protect people's privacy—including from repressive regimes or those that outlaw homosexuality.** These should be developed in conjunction with the Information Commissioner's Office and following consultation with groups including representatives of the LGBTQ+ community, victims of domestic abuse, journalists, and freedom of expression organisations. **Enforcement of people's data privacy and data rights would remain with the Information Commissioner's Office, with clarity on information sharing and responsibilities.**

Societal Harm and Safety by Design:

- Later in this report we discuss new offences proposed by the Law Commission around harm-based or knowingly false communications. These may be helpful in some instances in tackling disinformation, but they also have limitations. The harm-based offence relates specifically to psychological harm, so may not be applicable to vaccine disinformation, and knowingly false means just that—the person sending the communication must know it is untrue. **It is also unclear whether the latter offence would assist in cases of disinformation trying to disrupt elections, as the harm is based on psychological or physical harm, rather than harm to an institution, process, state, or society. The Elections Bill, which is currently making its way through Parliament with the intention "to strengthen the integrity of the electoral process" should address the issue of disinformation which aims to disrupt elections.**
- **Disinformation and misinformation surrounding elections are a risk to democracy. Disinformation which aims to disrupt elections must be addressed by legislation. If the Government decides that the Online Safety Bill is not the appropriate place to do so, then it should use the Elections Bill which is currently making its way through Parliament.**
- Disinformation which aims to disrupt elections must be addressed by legislation.

- **The Information Commissioner, Elizabeth Denham, has stated that the use of inferred data relating to users' special characteristics as defined in data protection legislation, including data relating to sexual orientation, and religious and political beliefs, would not be compliant with the law. This would include, for example, where a social media company has decided to allow users to be targeted with content based on their data special characteristics without their knowledge or consent. Data profiling plays an important part in building audiences for disinformation, but also has legitimate and valuable uses. Ofcom should consult with the Information Commissioner's Office to determine the best course of action to be taken to investigate this and make recommendations on its legality.**
- **We recommend content-neutral safety by design requirements, set out as minimum standards in mandatory codes of practice.** These will be a vital part of tackling regulated content and activity that creates a risk of societal harm, especially the spread of disinformation. **For example, we heard that a simple change, introducing more friction into sharing on Facebook, would have the same effect on the spread of mis- and disinformation as the entire third-party fact checking system.**
- **We also recommend far greater transparency around system design, and particularly automated content recommendation.** This will ensure the regulator and researchers can see what the platforms are doing, assess the impact it has and, in the case of users, make informed decisions about how they use platforms.

- **Fair Vote UK Analysis – Societal Harm, Anonymity and Platform Design:**
  - The renewed focus on activity rather than content is a welcome shift, reflecting the fact that moderating content - making the Online Safety Bill a so-called "take down" bill - is the wrong direction of travel, both threatening freedom of expression and failing to regulate the business model and incentive system that underpins content delivery.
  - The Safety by Design mitigation strategies are excellent, particularly: increasing transparency, countering algorithmic power, and de-virilizing and diversifying content through "friction" mitigation - a proven way to preserve free speech while limiting free reach of content that poses societal harm at scale.
  - The Joint Committee's assertion that targeted advertising using protected characteristics is likely already illegal - and noting the need for an immediate investigation is welcome and should not go unnoticed. This is a serious and necessary check on the power of micro-targeted ad-driven business models. Additional consideration should be made to the use of geographic targeting in political advertising as it relates to local vs. national spending limits.

- The recommendations around anonymity are also strong, striking the appropriate balance between protection of marginalised groups and those fearing persecution under repressive regimes and the right of a user to both verify their account and filter their online experience to exclude anonymous accounts if they so choose, helping to reduce online bullying and abuse. These are the so-called "right to verify" and "right to filter" provisions of which we are supportive. These recommendations, like most in this report, continue to take a welcome systems-focused approach that is essential in contending with foreign and domestic bot farms, anonymous abuse and anonymity-amplified disinformation campaigns.
- Ensuring that regulators have access to algorithms is both welcome and essential.
- **Important caveat:** The report claims that the Elections Bill should be used as the mechanism to address election and democracy-related disinformation. At present, the Elections Bill has no provisions related to this and would require robust amendments to do so. With the exception of the digital imprints regime, we believe election-related disinformation should be dealt with through the Online Safety Bill.

Section Four: Safety Duties Relating to Adults: Illegality and Legal but Harmful

Quotes and Paraphrases from the Report:

Illegal Content and Activity:

- We believe the scope of the Bill on illegal content is too dependent on the discretion of the Secretary of State. This downplays the fact that some content that creates a risk of harm online potentially amounts to criminal activity. The Government has said it is one of the key objectives of the Bill to remove this from the online world.
- **We recommend that criminal offences which can be committed online appear on the face of the Bill as illegal content.** This should include (but not be limited to) hate crime offences (including the offences of "stirring up" hatred), the offence of assisting or encouraging suicide, the new communications offences recommended by the Law Commission, offences relating to illegal, extreme pornography and, if agreed by Parliament, **election material that is disinformation about election administration, has been funded by a foreign organisation targeting voters in the UK or fails to comply with the requirement to include information about the promoter of that material in the Elections Bill.**

Reform of the Criminal Law:

- Implementation of the Law Commission's recommendations on reforming the Communications Offences and Hate Crime will allow the behaviour covered by the new offences to be deemed illegal content. We believe this is a significant enhancement of the protections in the Bill, both for users online but also for freedom of expression by introducing greater certainty as to content that online users should be deterred from sharing.
- **We endorse the Law Commission's recommendations for new criminal offences in its reports, Modernising Communications Offences and Hate Crime Laws.** The reports recommend the creation of new offences in relation to cyberflashing, the encouragement of serious self-harm, sending flashing images to people with photo-sensitive epilepsy with intent to induce a seizure, sending knowingly false communications which intentionally cause non-trivial emotional, psychological, or physical harm, communications which contain threats of serious harm and stirring up hatred on the grounds of sex or gender, and disability. We welcome the Secretary of State's intention to accept the Law Commission's recommendations on the Communications Offences. The creation of these new offences is absolutely essential to the effective system of online safety regulation which we propose in this report. We recommend that the Government bring in the Law Commission's proposed Communications and Hate Crime offences with the Online Safety Bill, if no faster legislative vehicle can be found. Specific concerns about the drafting of the offences can be addressed by Parliament during their passage.

Identifying illegal content:

- The criminal law is designed to establish whether or not an individual is guilty of an offence to a high standard of proof following an extensive, and adversarial, legal process. Since an individual's liberty and good name may be at stake, the criminal law requires that all elements of an offence be proved so that jurors are "satisfied so you are sure" or convinced "beyond reasonable doubt" before an offender is found guilty.
- The draft Bill addresses the problem of how some illegal content can be identified in practice by requiring Ofcom to publish a Code of Practice on terrorism content and CSEA content. It does not require such a Code of Practice for the wider duties around illegal content.
- **We recommend** that Ofcom be required to issue a binding Code of Practice to assist providers in identifying, reporting on and acting on illegal content, in addition to those on terrorism and child sexual exploitation and abuse content. As a public body, Ofcom's Code of Practice will need to comply with human rights legislation (currently being reviewed by the Government) and this will provide an additional safeguard for freedom of expression in how providers fulfil this requirement. With this additional safeguard, and others we discuss elsewhere in this report, **we consider that the test for illegal content**

in the Bill is compatible with an individual's right to free speech, given providers are required to apply the test in a proportionate manner that is set out in clear and accessible terms to users of the service.

- **We recommend** that the highest risk service providers are required to archive and securely store all evidence of removed content from online publication for a set period of time, unless to do so would in itself be unlawful. In the latter case, they should store records of having removed the content, its nature and any referrals made to law enforcement or the appropriate body.
- **We recommend that the Secretary of State's power to designate content relating to an offence as priority illegal content should be constrained.** Given that illegal content will in most cases already be defined by statute, this power should be restricted to exceptional circumstances, and only after consultation with the Joint Committee of Parliament that we recommend in Chapter 9, and implemented through the affirmative procedure. The Regulator should also be able to publish recommendations on the creation of new offences. We would expect the Government, in bringing forward future criminal offences, to consult with Ofcom and the Joint Committee as to whether they should be designated as priority illegal offences in the legislation that creates them

Legal but Harmful Content:

- One of the problems this legislation must grapple with is defining what creates a risk of harm to adults. Clause 11 attempts this in a broad way, and we have heard throughout our inquiry that this will make it difficult to apply, as well as open to legal challenge
- Clause 11 of the draft Bill has been widely criticised for its breadth and for delegating the authority of the state to service providers over the definition of content that is harmful and what they should do about it. We understand its aims and that the Government intended it primarily as a transparency measure over something companies are already doing. As drafted, however, it has profound implications for freedom of speech, is likely to be subject to legal challenge and yet may also allow companies to continue as they have been in failing to tackle online harm.
- **We recommend that Clause 11 of the draft Bill is removed. We recommend that it is replaced by a statutory requirement on providers to have in place proportionate systems and processes to identify and mitigate reasonably foreseeable risks of harm arising from regulated activities defined under the Bill.** These definitions should reference specific areas of law that are recognised in the offline world, or are specifically recognised as legitimate grounds for interference in freedom of expression. **For example, we envisage it would include:**
    - **Abuse, harassment or stirring up of violence or hatred based on the protected characteristics in the Equality Act 2010 or the characteristics for which hatred may be an aggravating factor under Crime and Disorder Act 1998 and section 66 of the Sentencing Act 2020;344**

- Content or activity likely to cause harm amounting to significant psychological distress to a likely audience (defined in line with the Law Commission offence);
- **Threatening communications that would lead a reasonable person to fear that the threat might be carried out;**
- Knowingly false communications likely to cause significant physical or psychological harm to a reasonable person;
- Unsolicited sending of pictures of genitalia;
- Disinformation that is likely to endanger public health (which may include anti-vaccination disinformation);
- Content and activity that promotes eating disorders and self-harm;
- **Disinformation that is likely to undermine the integrity and probity of electoral systems.**

- **We recommend that Ofcom be required to issue a mandatory code of practice to service providers on how they should comply with this duty.** In doing so they must identify features and processes that facilitate sharing and spread of material in these named areas and set out clear expectations of mitigation and management strategies that will form part of their risk assessment, moderation processes and transparency requirements. While the code may be informed by particular events and content, it should be focused on the systems and processes of the regulated service that facilitates or promotes such activity rather than any individual piece of content. **We envisage that this code would include (but not be limited to):**
  - **the moderation of user generated content to cover the use of AI for moderation;**
  - **the appropriate thresholds for human oversight;**
  - **the level of expertise needed for human moderation;**
  - **dedicated teams for election periods and involve relevant bodies–with planned circuit breakers;**
  - **the use of fact checking in proportion to reach and risk;**
  - **a transparency requirement on the top 20 viral messages, published on a monthly basis;**
  - **user control over their curation, including being joined to groups without permission; and**
  - **targeting through protected characteristics and or political affiliation.**
- **We recommend that additions to the list of content that is harmful should be by statutory instrument from the Secretary of State. The statutory instrument should be subject to approval by both Houses, following a report from the Joint Committee we propose in Chapter 9.** Ofcom, when making recommendations, will be required by its existing legal obligations to consider proportionality and freedom of speech rights. The Joint Committee should be specifically asked to report on whether the proposed addition is a justified interference with freedom of speech rights.

Accessibility/Consistency in Terms and Conditions:

- **<u>We recommend</u> that the Bill mandates service providers to produce and publish an Online Safety Policy, which is referenced in their terms and conditions, made accessible for existing users and made prominent in the registration process for new users.** This Online Safety Policy should: explain how content is promoted and recommended to users, remind users of the types of activity and content that can be illegal online and provide advice on what to do if targeted by content that may be criminal and/or in breach of the service providers' terms and conditions and other related guidelines.
    - **The Online Safety Policy should be produced in an accessible way and should be sent to all users at the point of sign up and, as good practice suggests, at relevant future points. "Accessible" should include accessible to children** (in line with the Children's Code), where service providers allow child users, and accessible to people with additional needs, including physical and learning disabilities. Ofcom should produce a Code of Practice for service providers about producing accessible and compliant online safety policies and on how they should make them available to users to read at appropriate intervals in line with best practice (for example, when the user is about to undertake an activity for the first time or change a safety-relevant setting).

- **Fair Vote UK Analysis – Illegality and Legal but Harmful:**
    - We support the Committee's recommendations on illegality and its emphasis on the Law Commission's findings as well. Requiring binding codes of practice for platforms to use to contend with illegal content and mandating archival of removed illegal content is sensible and proportionate.
    - Secretary of State powers related to "priority illegal content" should be constrained as the Committee suggests and the maintaining of a permanent joint committee offers a strong way of ensuring democratic accountability and the primacy of Parliament.
    - We accept that additions to the list of content that is harmful should be by statutory instrument from the Secretary of State *but only* if the statutory instrument is subject to approval by both Houses, following a report from the Joint Committee proposed in Chapter 9.
    - The replacement of Clause 11 with a statutory requirement to mitigate content likely to cause harm at scale is a welcome change. OfCom's mandatory code of conduct for these "proportionate systems and processes" to respond to harmful content are intended to ensure that the systems can rapidly identify content and contend with virality, which is the primary factor in determining likelihood to cause harm at scale. Questions remain as to exactly what this will look like and what kind of exemptions will remain in place and we look forward to a more detailed discussion on these points.

- Mandating the publication of an accessible Online Safety Policy is prudent for purposes of transparency and recenters platform design around safety instead of engagement and profitability.

Section Six: Scope of the Bill

Quotes and Paraphrases from the Report:

Categorisation of services

- **We recommend that the categorisation of services in the draft Bill be overhauled. It should adopt a more nuanced approach, based not just on size and high-level functionality, but factors such as risk, reach, user base, safety performance, and business model.** The draft Bill already has a mechanism to do this: the risk profiles that Ofcom is required to draw up. We make recommendations in Chapter 8 about how the role of the risk profiles could be enhanced. We recommend that the risk profiles replace the "categories" in the Bill as the main way to determine the statutory requirements that will fall on different online services. This will ensure that small, but high risk, services are appropriately regulated; whilst guaranteeing that low risk services, large or small, are not subject to unnecessary regulatory requirements.

Paid for Advertising Exclusion

- **We recommend that clause 39(2) is amended to remove "(d) paid-for advertisements" to bring such adverts into scope. Clause 39(7) and clause 134(5) would therefore also have to be removed.**
- **We recommend that the Bill make clear Ofcom's role will be to enforce the safety duties on providers covered by the online safety regulation, not regulate the day-to-day content of adverts or the actions of advertisers.**

- **Fair Vote UK Analysis – Scope of the Bill:**
    - Risk is a much better metric than size for moderation and we are supportive of the new categorisation suggestions.
    - We are extremely supportive of paid-for ads being brought into scope. It is an affront to equal protection of freedom of expression that content is exempt if it has been purchased.

Section Seven: Freedom of Speech, Journalism and Democratic Content

FAIR VOTE

Quotes and Paraphrases from the Report:

Freedom of Expression:

- **We propose a series of recommendations throughout this report to strengthen protection for freedom of expression. These include greater independence for Ofcom, routes for individual redress beyond service providers, tighter definitions around content that creates a risk of harm, a greater emphasis on safety by design, a broader requirement to be consistent in the applications of terms of service, stronger minimum standards and mandatory codes of practice set by Ofcom (who are required to be compliant with human rights law), and stronger protections for news publisher content. We believe these will be more effective than adjustments to the wording of Clause 12.**

Journalistic and Democratic Content Exemptions:

- **We recommend that the news publisher content exemption is strengthened** to include a requirement that news publisher content should not be moderated, restricted or removed unless it is content the publication of which clearly constitutes a criminal offence, or which has been found to be unlawful by order of a court within the appropriate jurisdiction. **We recommend that the Government look at how bad actors can be excluded from the concept of news publisher.** We suggest that they may wish to exclude those that have been repeatedly found to be in breach of The Ofcom Broadcasting Code, or are publications owned by foreign Governments. Ofcom should also examine the use of new or existing registers of publishers. We are concerned that some consumer and business magazines, and academic journals, may not be covered by the Clause 40 exemptions. **We recommend that the Department consult with the relevant industry bodies** to see how the exemption might be amended to cover this off, without creating loopholes in the legislation.
    - Our recommendations to narrowly define content that is harmful to adults by way of reference to existing law should provide some of the extra clarity service providers need to help protect freedom of expression. At the same time, journalism and content of democratic importance have long been recognised as vital in a democratic society and should be given specific consideration and protection by providers, who have significant influence over the information we see. We have heard concerns around the definitions used however, and about the ability of the providers to interpret and apply them consistently. **We feel that "democratic importance" may be both too broad—creating a loophole to be exploited by bad actors—and too narrow—excluding large parts of civil society.** Similarly, we are concerned that any definition of journalistic content that is designed to capture citizen journalism would be so broad it would render the consistent application of the requirement almost impossible, and see the

expedited complaints route overwhelmed by people claiming without merit to be journalists in order to have their content reinstated. “Public interest” might be more useful in ensuring that content and activity is judged on its merit, rather than its author.

- **We recommend that the existing protections around journalistic content and content of democratic importance should be replaced by a single statutory requirement to have proportionate systems and process to protect ‘content where there are reasonable grounds to believe it will be in the public interest’.** Examples of content that would be likely to be in the public interest would be journalistic content, contributions to political or societal debate and whistleblowing. **Ofcom should produce a binding Code of Practice on steps to be taken to protect such content and guidance on what is likely to be in the public interest, based on their existing experience and case law.** This should include guidance on how appeals can be swiftly and fairly considered. Ofcom should provide guidance to companies in cases of systemic, unjustified take down of content that is likely to be in the public interest. This would amount to a failure to safeguard freedom of expression as required by the objectives of the legislation.

- **Fair Vote UK Analysis – Freedom of Expression, Journalistic and Democratic Speech Exemption:**
  - This report wisely views freedom of expression in the context of systems and algorithms deployed by digital platforms, acknowledging that platforms have significant control over the content that we see and equally, who sees the content that we produce. Stronger protections for news publisher content is troubling without a clear definition of what a news publisher is and we look forward to greater clarity around this point.
  - The draft bill leaves a lot of unanswered questions around how the exemptions for journalistic and democratically important content will function. The recommendation that the “Government look at how bad actors can be excluded from the concept of the news publisher” is therefore critical and we look forward to being part of this process.
  - We agree that “democratic importance” is far too broad a term for bad actors and too restrictive for civil society. New statutory codes replacing protections for journalistic and democratic content with content reasonably presumed to be in the “Public Interest” is a clearer way forward. The recommendations they propose are sensible and we broadly support them, depending on what is included in the binding Code of Practice produced by Ofcom.

Section Eight: Role of the Regulator

FAIR VOTE

Quotes and Paraphrases from the Report:

Risk Assessments:

- **The Bill's provision that Ofcom should develop risk profiles based on the characteristics of services should be strengthened. Ofcom should begin drawing up risk profiles immediately so that they are ready to be actioned when the Bill becomes law.** Risk profiles should reflect differences in the characteristics of the service. These could include (but are not limited to) risks created by algorithms; risks created by a reliance on artificial intelligence moderation; risks created by unlimited 'one-click' sharing; risks caused by "engagement" maximising design features; risk of unsupervised contact between adults and children which may give rise to grooming; risks caused by surveillance advertising; and such other risks as Ofcom identifies in its overall risk assessment, as well as platform design, risk level, end-to-end encryption, algorithmic design, safety by design measures, and the service's business model and overall corporate aim. Ofcom should also be able to take into account whether a company has been the subject of a super complaint, other legal proceedings or publicly documented evidence of poor performance e.g. independent research, a poor monitoring report in the EU's Code of Conduct for Illegal Hate, or whistleblowers' evidence.
- **Ofcom should be required to set binding minimum standards for the accuracy and completeness of risk assessments.** Ofcom must be able to require a provider who returns a poor or incomplete risk assessment to redo that risk assessment. Risk assessments should be carried out by service providers as a response to the Online Safety Act before new products and services are rolled out, during the design process of new features, and kept up to date as they are implemented.
- **The Bill should be amended to clarify that risk assessments should be directed to "reasonably foreseeable" risks**, to allow Ofcom greater leeway to take enforcement action against a company that conducts an inadequate risk assessment.

Auditing:

- In bringing forward the final Bill, **we recommend the Government publish an assessment of the audit powers given to Ofcom and a comparison to those held by the Information Commissioner's Office and the Financial Conduct Authority.** Parliament should be reassured that the Bill will give Ofcom a suite of powers to match those of similar regulators. Within six months of the Act becoming law, Ofcom should report to Parliament on how it has used those powers.
- **We recommend that the largest and highest-risk providers should be placed under a statutory responsibility to commission annual, independent third-party audits of the effects of their algorithms, and of their risk assessments and transparency reports.** Ofcom should be given the explicit power to review these and undertake its own audit of these or any other regulated service when it feels it is required. Ofcom should develop a framework for the effective regulation of algorithms based on the requirement for, and auditing of, risk assessments.

FAIR VOTE

Coregulation:

- In taking on its responsibilities under the Bill, Ofcom will be working with a network of other regulators and third parties already working in the digital world. **We recommend that the Bill provide a framework for how these bodies will work together including when and how they will share powers, take joint action, and conduct joint investigations.**
- We reiterate the recommendations by the House of Lords Communications and Digital Committee in their Digital Regulation report: that **regulators in the Digital Regulation Cooperation Forum should be under a statutory requirement to cooperate and consult with one another, such that they must respect one another's objectives, share information, share powers, take joint action, and conduct joint investigations**; and that to further support coordination and cooperation between digital regulators including Ofcom, the Digital Regulation Cooperation Forum should be placed on a statutory footing with the power to resolve conflicts by directing its members.
- The draft Bill does not give Ofcom co-designatory powers. Ofcom is confident that it will be able to co-designate through other means. **The Government must ensure that Ofcom has the power to co-designate efficiently and effectively, and if it does not, this power should be established on the face of the Bill.**

Codes of Practice:

- **The Bill should be amended to make clear that Codes of Practice should be binding on providers.** Any flexibility should be entirely in the hands of and at the discretion of the Regulator, which should have the power to set minimum standards expected of providers. They should be subject to affirmative procedure in all cases.

Criminal Liability:

- **The Bill should require that companies' risk assessments be reported at Board level, to ensure that senior management know and can be held accountable for the risks present on the service, and the actions being taken to mitigate those risks.**
- **We recommend that a senior manager at board level or reporting to the board should be designated the "Safety Controller" and made liable for a new offence: the failure to comply with their obligations as regulated service providers when there is clear evidence of repeated and systemic failings that result in a significant risk of serious harm to users.** We believe that this would be a proportionate last resort for the Regulator. Like any offence, it should only be initiated and provable at the end of an exhaustive legal process.
- The Committee welcomes the Secretary of State's commitment to introduce criminal liability within three to six months of Royal Assent and strongly recommends that criminal sanctions for failures to comply with information notices are introduced within three months of Royal Assent.

Secretary of State Powers:

- **The power for the Secretary of State to exempt services from regulation should be clarified to ensure that it does not apply to individual services.**

- **The powers for the Secretary of State to a) modify Codes of Practice to reflect Government policy and b) give guidance to Ofcom give too much power to interfere in Ofcom's independence and should be removed.**

Media Literacy:

- If the Government wishes to improve the UK's media literacy to reduce online harms, **there must be provisions in the Bill to ensure media literacy initiatives are of a high standard. The Bill should empower Ofcom to set minimum standards for media literacy initiatives** that both guide providers and ensure the information they are disseminating aligns with the goal of reducing online harm.

- **Fair Vote UK Analysis – Role of the Regulator:**
    - We welcome all of the recommendations for risk assessments, auditing, codes of practice and criminal liability.
    - It is crucial to curb Secretary of State powers, as suggested by the Committee. It is of significant democratic concern when a party-political minister of state has unilateral sway over matters of such vast importance.
    - We agree with media literacy recommendations as they form the bedrock of generational change in this topic area. They must be backed by a strategy to engage educators, civil society and academics in the development of best-practice curriculum as well as a reasonable budget to ensure adequate delivery of programs is possible.

Section Nine: Transparency and Oversight:

Quotes and Paraphrases from the Report:

Transparency for users:

- **We recommend that Ofcom specify that transparency reports produced by service providers should be published in full in a publicly accessible place.** Transparency reports should be written clearly and accessibly so that users and prospective users of the service can understand them, including children (where they are allowed to use the service) and disabled people.
- **We recommend that the Bill require transparency reporting on a regular, proportionate basis, with the aim of working towards standardised reporting as the regulatory regime matures.** The Bill should require minimum standards of accuracy and transparency about how the report was arrived at and the methodology used in research. For providers of the highest risk services, the outcome of the annual audits recommended in paragraph 340 should be required to be included in the transparency report.
- We agree with the list of information that Ofcom can require as part of its transparency reporting powers and recommend that it should have the clear power to request any other information. We recommend that transparency reporting should aim to create a

competitive marketplace in respect of safety, where people can reasonably compare, using robust and comparable information, performance of services as they operate for UK users. **We suggest Ofcom also be able to require information be published in transparency reports including (but not limited to):**

- **a)Safety by design features;**
- **b)Most viewed/engaged with content by month;**
- **c)Most recommended content by month by age group and other demographic information (where that information is collected);**
- **d)Their terms and conditions;**
- **e)Proportion of users who are children;**
- **f)Proportion of anonymous users;**
- **g)Proportion of content breaching terms and conditions;**
- **h)Proportion of content breaching terms and conditions removed;**
- **i)Proportion of appeals against removal upheld;**
- **j)Proportion of appeals against removal, by both recognised news publishers and other users on the grounds of public interest, upheld; and**
- **k)Time taken to deal with reports.**

Joint Committee on Online Regulation:

- ***We agree with other Committees that it is imperative that digital regulation be subject to dedicated parliamentary oversight. To achieve this, we recommend a Joint Committee of both Houses to oversee digital regulation*** *with five primary functions: scrutinising digital regulators and overseeing the regulatory landscape, including the Digital Regulation Cooperation Forum; scrutinising the Secretary of State's work into digital regulation; reviewing the codes of practice laid by Ofcom any legislation relevant to digital regulation (including secondary legislation under the Online Safety Act); considering any relevant new developments such as the creation of new technologies and the publication of independent research or whistleblower testimonies; and helping to generate solutions to ongoing issues in digital regulation.*

- **Fair Vote UK Analysis – Transparency and Oversight:**
  - Transparency is crucial. For regulators to mandate and subsequently maintain safety by design principles, the mechanics of the algorithms and targeting systems must be open to scrutiny by independent researchers, government bodies and the general public. The committee's suite of recommendations are reasonable and commensurate.
  - Similarly, dedicated, permanent Parliamentary oversight ensures cross-party democratic accountability where Parliament maintains primacy, ensuring that no single government can unduly influence the online information ecosystem.

Section Eleven: Conclusion

The report states:

This Report must be understood as a whole document, comprising a cohesive set of recommendations working in tandem to produce a new vision of the Online Safety Act. The Government should not seek to isolate single recommendations without understanding how they fit into the wider manifesto laid out by the Committee. Taken as a whole, our recommendations will ensure that the Bill holds platforms to account for the risks of harm which arise on them and will achieve the Government's ultimate aim of making the United Kingdom the safest place in the world to be online.

- **Fair Vote UK analysis: Conclusion:**
    - **We concur with the Committee's that this bill must be viewed in a context of interrelated interventions geared towards an overarching purpose of creating a safer digital environment that is systematically harm-reducing. We appreciate the Committee's systems-focused approach and recommendations and encourage the Government to adopt the full, cohesive set of recommendations.**