17 May 2019

Director, Online Content and eSafety Section
Department of Communications and the Arts
GPO Box 2154
CANBERRA ACT 2601

By Email: onlinesafety@communications.gov.au

Dear Director,

Re: Online Safety Charter consultation

Electronic Frontiers Australia (EFA) appreciates the opportunity to provide this submission in relation to the Online Safety Charter consultation. EFA's submission is contained in the following pages. This submission is made within the extended timeframe approved via email received from the Department of Communications and the Arts.

About EFA

Established in January 1994, EFA is a national, membership-based non-profit organisation representing Internet users concerned with digital freedoms and rights. EFA is independent of government and commerce, and is funded by membership subscriptions and donations from individuals and organisations with an altruistic interest in promoting civil liberties in the digital context. EFA members and supporters come from all parts of Australia and from diverse backgrounds.

Our major objectives are to protect and promote the civil liberties of users of digital communications systems (such as the Internet) and of those affected by their use and to educate the community at large about the social, political and civil liberties issues involved in the use of digital communications systems.

Yours sincerely,

Angus Murray
Chair of the Policy Committee
Electronic Frontiers Australia

Executive Summary

It is prudent to note that technology is neither good nor bad, it is merely a tool.

We note that this charter places onus on technology firms, however there needs to be a balance where users themselves need to be held responsible for the content that they may create or share when using these services. In situations where users are also content creators, it should be noted the Australian public should also be required to have regard to Australian law when publishing content.

Even though the terms are used repeatedly in the charter, no definitions for the terms “harmful”, promoting “self harm or criminal activity”, or “inappropriate” or examples of such content is offered in the charter. It’s difficult for EFA to respond to a call for consultation, and provide feedback for a proportional response, when key terms are not defined. (Appendix B offers an example of “illegal” content only). It is essential that these terms be fully defined, as they are open to vast differences in interpretation (i.e. “harmful” to whom? Is harm defined by physical damage, financial, emotional? Who makes this determination? What is “inappropriate”?).

Furthermore, there seems to be scant consideration in the charter consultation for cross jurisdictional issues: These issues include questions where providers may be in the position of contravening laws of other jurisdictions to meet the requirements of the charter, and content providers making the decision to cease operation within Australia, simply because the charter is too onerous compared to operating in states that do not have such onerous compliance.

Introduction

Scope

"The draft Charter is directed towards technology firms that offer the opportunity for users in Australia to interact or connect and technology firms whose services and products enable users to access content and information."

The scope statement in the proposed Charter has been designed to be deliberately broad and to identify all firms with a possible role in promoting online safety. All such firms are referred to collectively in subsequent clauses as 'technology firms'. The charter needs to allow a reader to understand the different obligations which may apply to different firms because it is not realistic to expect that all 'technology firms' can or should be regulated in the same manner. We are concerned about the practical ability of firms to adhere to a Charter where they are:

- Not for profit,
- Overseas based,
- Volunteer run,
- Small,
- Purely a carriage service (eg a VPN provider) or,
- Largely serving users overseas.

The human resources required to implement the charter may be prohibitive for non-profit or charity status technology firms. As such, this charter may result in a number of such firms choosing not to operate in Australia - which may be counter to the charter's objectives. (Such as those with the intent to provide support services to the vulnerable communities discussed in the charter such as online children's mental health support forums.)

Limits on the reach of the charter are not stated, e.g. should extended to technology firms that provide software for other technology firms? Is the intention that existing content would be required to be retroactively reviewed for its alignment to Australian law? (Wikipedia contains over 5,835,000 articles in English alone)

The scope of the charter makes no mention of legal jurisdiction in which the technology firm operates. There are likely legal limits to which jurisdictions it can or should apply.

It may not be possible for 'technology firms' that are unable to comply with the charter to effectively exclude their services from Australian users.

For example, it may be possible that a foreign organisation with no understanding of Australian law, or intentional marketing in Australia find themselves in breach of the charter whilst complying with the highest standards required in their legal jurisdiction.

The scope of the charter should be limited to for-profit commercial organisations deliberately targeting the mentioned Australian audience.

"While the proposed scope of the Charter is broad, it is acknowledged that the digital media landscape is not homogeneous, and that not all technology firms should be expected to demonstrate, or implement, identical measures in relation to online safety."
It is acknowledged that the charter is not intended to apply uniformly to all firms and their services, however this creates potential points of uncertainty, with 'technology firms' unsure of their obligations under the draft charter. Appropriate points of divergence should be articulated such that firms are better able to determine how to align with the charter (i.e. is it the intention that this divergence would be based on type of content (i.e. news vs social interaction), number or geography of user base (i.e. millions of users worldwide vs two thousand users in Victoria, AUS), drivers (i.e. charity, non-profit or commercial) or a range of other criteria.

Test case: Wikipedia (foreign non-profit)
"Wikipedia is a multilingual, web-based, free encyclopedia that is based on a model of openly editable content. It is the largest and most popular general reference work on the Internet, and is named as one of the most popular websites. It is owned and supported by the Wikimedia Foundation, a non-profit organization which operates on money it receives from its annual fund drives." (reference Wikipedia Foundation, 2019; Wikepedia; Available:https://en.wikipedia.org/wiki/Wikipedia April 7, 2019).

As stated above, Wikipedia is a web-based encyclopedia containing content that is both **produced** and **consumed** by its users. This dual usage model is important to consider in reference to the charter.

Attachment A

1.1 Content identification

"Technological solutions should be fully utilised by technology firms to identify illegal and harmful content"
The charter should set out the details of such currently available technological solutions that achieve this objective to the Government's satisfaction such that firms can seek to apply them. Clarity should be provided on the wording "fully utilised", as it is ambiguous in its meaning (i.e. does this mean applied to all content, content produced by all users? Could private voice calls between two parties be considered as content under the proposed charter?)

Further, given the wording "technological solutions" - Is it the intention that human content moderators would not be sufficient if "technological solutions" are not feasible or available?

The proposed Charter does not include a definition of 'harmful' that would enable a user of the document to determine the intended meaning. If the intent is to include a requirement to remove content that is not illegal but that is harmful then careful definition would be needed to allow that. EFA is not aware of any such broadly accepted definition.

"There should be a specific point of contact within each technology firm for the referral of complaints about illegal and harmful content or legal notices from Australian authorities. This point of contact should be equipped and trained to manage Australian referrals, with a good understanding of relevant Australian legal requirements."
As the scope is currently written in the proposed Charter not all 'technology firms' will be of a scale where it is reasonable to expect them to have a capacity to respond immediately, or even quickly, to complaints. In a similar way not all firms will have access to 'training' or 'a good knowledge of Australian law'.

Given the ambiguity of the terminology in this Charter, it should set out the details of training available to firms to identify content that is "illegal" under Australian law, "harmful", promoting "self harm or criminal activity", or "inappropriate" such that they can meet their obligations under the charter. The charter should also detail currently available complaints handling services that achieve the objective as set out in this section of the charter to the Government's satisfaction such that firms can seek to either emulate them or obtain their services.

In the case of Wikipedia, under this requirement, the non-profit Wikipedia foundation would be required to employ someone with a 'good legal understanding of relevant Australian legal requirements' in order to provide their essential learning platform in Australia. Further there is not currently any practical way of restricting Australians from accessing or contributing to Wikipedia if the Wikipedia foundation cannot comply with this requirement.

1.2 Content moderation

"The systems employed by technology firms should have the capability and capacity to moderate illegal and harmful content.
Further detail is required to ensure firms have clear direction on how to meet this requirement. For example, is it sufficient that a firm has a human moderator who has the ability to moderate content?
Additionally, clarification on the applicability of this charter is critical to the understanding of this requirement. For example, does the charter apply to Australian technological firms only? Is moderation only required for Australian users? Is moderation only required for content made available to Australian users?
Is it the intention of the Government that firms ringfence content for Australian audiences and only provide that is "moderated" in line with this requirement? If this were possible, this would effectively allow firms to decide the content is available to Australians - which seems to be at

odds with the intention of the charter. This is likely to have a chilling effect on firms making content *available* to Australian audiences - and in accepting content *created* by Australians (in the example of Wikipedia).

“Where feasible, this should include a triaging system to ensure high risk content (e.g. content promoting self-harm or criminal activity) is addressed expeditiously and lower risk content is reviewed and actioned within a longer period (for example, within 24 hours).”

This requirement expands the previously stated obligations to impose the additional necessity for technology firms to define content that also ***promotes “self-harm or criminal activity”*** in addition to content that is “illegal or harmful”. The concept of “self-harm” in particular is hugely subjective. Without detailed guidance it is not feasible to expect technology firms to be able to make such determinations. It is heavily recommended that this requirement be amended to refer to actual Australian legal requirements to remove such requirements for subjectivity.

“Where appropriate, illegal, harmful or inappropriate content targeted towards a child should be removed immediately, and only reinstated once the complaint has been investigated and only if the complaint is not upheld.”

Similarly to the statement above, this requirement expands the obligations above to now impose the necessity for technology firms to define and identify “*inappropriate*” content as well as content that “self-harm or criminal activity”, and that is “illegal or harmful”. Again, without detailed guidance it is not feasible to expect technology firms to be able to make such determinations. It is heavily recommended that this requirement be amended to refer to actual Australian legal requirements to remove such requirements for subjectivity.

Further, this requirement necessitates that a determination be made as to the” legality”, “harmfulness”, “inappropriateness” and intended audience of content *before* a complaint has been reviewed - so that it can be “removed immediately”. This is requires expert judgement on the part of the reviewer, and clearly is not feasible to carry out such a detailed review “immediately”.

This requirement that a firm must act before considering the merits of a complaint opens risk of malicious or vexatious abuse of these provisions with potential consequences of legal disputation or financial loss. It is suggested that some assessment of complaints is always allowed before action is required so that frivolous, vexatious or malicious complaints can be set aside.

"This triaging system should ensure that complaints made by children, or by adults on behalf of children, are also expedited"

This additional requirement is unnecessary because the requirements above to remove inappropriate, illegal or harmful material targeting a child already demand prompt action. The age of the complainant is not relevant.

Under this requirement, firms would be required to collect and retain personal information on individuals making complaints such that it is possible to identify if the complainant is a child or an adult complaining on behalf of a child. This would require firms to store detailed personal information about children which would not have previously been necessary. This would then require firms to comply with further privacy obligations.

"The resources devoted to content moderation should be proportionate to the volume of content available to users and relevant to the Australian context. Human content moderators should meet minimum training standards."
As mentioned above, this requirement seems to require firms to restrict the content available to Australian audiences and restrict the content accepted from Australian contributors - to ensure only “moderated” content is available or used.

Wikipedia is likely the largest repository of knowledge (content) that has ever existed and as such would be expected to devote more resources to moderation than any other 'technology firm', even for-profits.

An enormous amount of human effort goes into content on Wikipedia, it is predominantly controlled (and moderated) by its users and many are anonymous. Wikipedia may need to limit contributions to registered Australian users who have been shown to 'meet minimum training standards' whilst still devoting an enormous amount of resources to content moderation. This requirement for “proportionate” additional content moderation to be provided by the Wikipedia Foundation would be prohibitive for Wikipedia to be available to Australian users.

1.3 Content removal
"Content that is clearly and unambiguously illegal under Australian law should be removed proactively by technology firms."
The charter sets requirements for firms to identify content that is “illegal” (under Australian law), “harmful”, promoting “self harm or criminal activity”, “inappropriate”. However, this requirement to proactively remove content is only applied to content that is “clearly and unambiguously illegal”. Clarification should be provided as to the charter’s expectations for content that is not in this category but is “harmful”, promoting “self harm or criminal activity”, “inappropriate”.

In the example of Wikipedia, the Wikipedia foundation would be required to obtain a detailed understanding of Australian law and what content would be considered illegal under Australian Law. This would need to be provided to its user moderators in order for them to enforce this requirement. In situations of ambiguity or perception, detailed legal guidance may be required to

determine the standing of such content. Clarity is required as to who is responsible for obtaining such legal guidance, the content creator or the technology firm moderator.

"Technology firms should take steps to prevent the reappearance of illegal, harmful or offensive content that has been removed."
The charter would benefit from more detail on the intended result of this statement in order to implement this recommendation. Once placed online, content is used, reused, copied and cached on multiple servers and sites worldwide. While this requirement in the charter may intend to remove content - this is unlikely to result in the content being "removed from the internet" altogether.

2.1 User behaviour
Clear minimum standards for online behaviour should be set and applied consistently across services and service providers.
The charter does not identify who is responsible for setting these "clear minimum standards" "across services and service providers" and ensuring they are "applied consistently". This needs to be clearly stated such that firms can comply.

2.2. User behaviour

"Banned users should not be able to open a new account in a different name or register a different user name."
This is extremely difficult (if not impossible, on current platforms) to implement. Tools to achieve this end are easily circumvented.

2.3 Account control

"Users should be able to freeze their account in real time."
This statement is unclear - and further detail is required as to what "freeze" is intended to mean. As it stands, it is not possible to determine if it is feasible for technology firms to comply with this requirement. Are "frozen" users unable to view any further content on a site or platform? Or are other site users unable to view content about "frozen users"? What is the charter's intention for this statement?

"Users under 16 years should be required to secure parental or guardian consent to open an account or register as a user. Verifying parental consent should require more than just ticking a box."
This proposal is out of step with global norms which allow users to open social media (Facebook, Google and Instagram) accounts at age 13 in most jurisdictions. Many schools register primary aged children with 'technology firms' without specific parental consent (e.g. Mathletics https://au.mathletics.com).

This requirement necessitates that all Australian users are required to register, and prove their age - with further identifying information required in the case of children.

Further guidance is needed on the minimum standards to meet this requirement for verifying parental consent. The charter clearly states that this cannot be “just ticking a box” but does not provide any further minimum standards. For example, typing “YES” be considered sufficient? Are parents details required to be stored by the firm?

“Parental control settings should be easy to use and difficult to circumvent.”
Clarity should also be provided as to the intended meaning of the terms “easy” and “difficult” with regards to these parental controls, such that firms can meet the charter’s standards.

“Users should be given full control of content safety options, such as the ability to delete unwanted comments, easily remove content...”
This requirement potentially stifles free speech and the ability to criticise online. Clarity is needed on the appropriate limits around a user’s control over comments and content. A user having agency over moderating, say, comments on their own Instagram profile/content seems appropriate, but allowing any user to edit or delete any comment on any part of, say, a typical online discussion forum takes things to another level - probably inappropriate.

Attachment B - Discussion Questions

1.1 Content identification

1. What are the examples of technology-facilitated solutions to enhance online safety, and how effective have these solutions been in addressing harms and mitigating risks?
No response.

2. What tools are available and have been deployed to address safety issues for live-streamed content as it occurs?
No response.

3. What is the best way to establish a single 24/7 contact point for Australian authorities to ensure there is a timely response?
EFA suggests that the Government provides a central contact point for Australians to make such complaints which can then be channeled to the correct firms. In a number of situations, it will be difficult for the average person to determine who to contact with their complaint, their ISP, the news provider, the social media service used to publicise it or the person commenting about it. As such, it is more appropriate for a central service to set rules for how to make this determination and then follow up on them.

1.2 Content moderation

4. Are there positive examples of flagging and content moderation? What makes these moderation systems work effectively and are they applicable to other services and Applications?

The charter states "Technology firms should keep a record of material that is taken down, and removed content should be preserved so that it is available if needed as evidence by Australian authorities", EFA strongly opposes the notion of technology firms being involved in forensic preservation for law enforcement, and questions whether material collected by a technology firm could meet chain of custody or evidentiary requirements. Removal of content, for the aims of preventing harms, should be the charter's limit of obligations for technology firms.

5. Is there an acceptable error rate for inappropriately flagged or misidentified content?
No response.

6. What is an appropriate time frame for moderation and removal of content?
No response.

7. How should content moderators be trained? What minimum standards should apply?
Given that this training will require detailed knowledge of Australian law, the Australian Government should provide this training to an agreed acceptable standard as required for firms to implement.

8. What sort of guidance should be available to moderators about dealing with vulnerable groups, such as children and Indigenous Australians?
Given that this training will require detailed knowledge of Australian law and social structures, the Australian Government should provide this training to an acceptable standard as required for firms to implement.

1.3 Content removal

9. Are there positive examples of identification and content removal practices? What makes these practices effective and appropriate?
No response.

10. How should records of removed content be kept to ensure that evidence is available if needed by authorities?
Given that this would involve firms retaining records of potentially incriminating content, the Australian Government or responsible enforcement agencies should be responsible for retaining this content.

11. Are there minimum requirements to uniquely identify content (for example, IP addresses of upload/posting source, geographic identifiers etc)? If so, please provide details.

Measures to uniquely identify images and video are easily subverted and are impractical. Such attempted measures would involve tracking individuals, their devices and their user identities on various services. This would certainly constitute an invasion of privacy for the vast majority of Australian users who are not undertaking the illegal activity concerned.

There is significant complexity in identifying content and content producers, especially in cross-jurisdictional situations, such as Europe's GDPR. There is a requirement for measures to be in place to ensure that content providers are not breaching laws of other jurisdictions, purely to satisfy the charter.

12. Can content be made invisible on a permanent basis? If so, how?

This would be impractical, and defeat the greater utility provided by sources such as the Internet Archive's Wayback Machine (https://www.archive.org).

13. Are there barriers to sharing of information about offensive content removed by an industry participant to prevent it being uploaded to another platform or distributed using another service?

No response.

14. What are the potential pitfalls and risks with content removal? How can these risks be mitigated?

No response

2.1 User behaviour

15. What should minimum standards of behaviour be? Should they be higher for products and services directed at children, or that have a substantial number of child users?

Standards for children should be set at the same current minimum standards as children's television or reading material. Further, minimum standards of behaviour should extend beyond children to other vulnerable groups.

16. How frequently should users be required to 'accept' or re-acknowledge terms of use, standards and policies?

No response

17. How should users be required to verify acceptance of terms of use, standards and policies?

No response

18. Are there positive examples of improving user experience currently in use?
No response

2.2 User support
19. Are there positive examples of user support systems and processes currently in use? What are the factors and characteristics of these systems and processes that make them effective?
No response

20. What timeframe is reasonable to respond to complaints and reports?
This will surely change drastically based on size, scale and resources of a service and the degree to which a firm is complying with this charter.

21. Should reporting and complaint response timeframes vary depending on the complainant (e.g. child or adult), the type of content or other factors?
As above, this could potentially require greater resources (such as AI or algorithms) not available to all firms and as such the answer may be different based on the firm concerned.

2.3 Account and device control
No response

22. What options are there for verifying age or ensuring that parental/guardian consent is provided? Is there an optimal method or methods?
Current measures to verify age or parent consent are easily subverted. It is unlikely that future verifications will be foolproof and may impose unnecessary complication and privacy concerns for valid users. Options for age verification should not lock technology firms into using proprietary software or systems. Proprietary systems may make mandatory verification financially impossible for non-profit organisations.

EFA understands that there is an age-verification system being employed within the United Kingdom, despite multiple setbacks in its implementation. EFA does not support age verification by third party, commercial, organisations.

Future verification systems should exist in the public sphere, ideally implemented in open source software, and be governed by clear and open privacy and security systems.

23. Are there positive examples of parental settings currently in use?
No response

24. Are there barriers to obtaining or using parental controls? How can these barriers be managed and overcome?
No response

2.4 Content management

No response to discussion questions in section 2.4