



Best Practice for User Comment Moderation

**Including commentary for organisations using
social media platforms**

July 2013

Table of Contents

| | |
|--|----|
| Introduction | 2 |
| Background | 3 |
| Summary of recommendations | 5 |
| Why enter this conversation? | 6 |
| People and organisations are using social media for different purposes | 7 |
| Legal treatment of user comments | 8 |
| The role of moderation on social media platforms | 9 |
| Tools made available by social media platforms | 12 |
| Appendix A: Table outlining different types of moderation | 13 |
| Appendix B: Developing a moderation policy | 15 |
| Appendix C: ACCC Guidance | 16 |
| Appendix D: Facebook tools | 19 |
| Appendix E: YouTube tools | 20 |
| Appendix F: Google+ tools | 21 |
| Contributors and Supporters | 22 |

Introduction

These guidelines and accompanying commentary set out recommendations for how organisations should moderate user generated comments that are posted to social media channels. It highlights the tools made available by the social media platform operators, identifies good business practice with respect to moderating user comments and general guidance around how to manage these conversations.

The signatories to this guideline all operate within Australia and represent social media platform operators and organisations who use these platforms to engage, as well as agencies who support their clients' efforts in this space. What we all have in common is our commitment to the future development of social media as a critical space within which these many different conversations can take place in an open and responsible manner.

The Interactive Advertising Bureau (IAB) is the peak trade association for online advertising in Australia and was incorporated in July 2010. IAB Australia's board includes representatives of AIMIA, Carsales.com, Fairfax Media, Google, Mi9, Network Ten, News Australia, REA Group, SBS, Telstra Media, TressCox Lawyers and Yahoo!7.

We recognise that there are many stakeholders in this conversation – large advertisers, small organisations, nonprofits, political parties, start-ups and individual Australians. All of us use and stand to benefit from social media and we welcome the contributions recently made to this conversation by the Australian Competition and Consumer Commission (ACCC)¹ and the New Zealand Advertising Standards Association². This guideline is very much designed to build upon the guidance issued by the ACCC and our hope in

¹ See appendix A

² www.asa.co.nz

putting this guideline out for consultation and discussion is that everyone who is potentially impacted by it can have their views heard.

Background

Australians' have embraced social media as a way of expressing their views, connecting and sharing with friends and family, and engaging with organisations and brands in new and powerful ways. 97% of online Australians use social media³, many of which are accessing these sites and services multiple times a day (e.g. 50% of Facebook users return daily). There are over 3,200,000,000 likes and comments on Facebook every day; that's 3.2 billion or 37,000 every second⁴. 100 million people take a social action (likes, shares, comments, etc.) on YouTube every week⁵. More than 50% of videos on YouTube have been rated or include comments from the community⁶.

Australian organisations have also embraced the opportunity of reaching so many of their existing and potential customers through these platforms. 30% of small to medium enterprises (SMEs) and 47% of medium enterprises use social media in their business⁷, five in ten SMEs report that using social media has had a positive impact on their business⁸. The ways in which organisations can engage with their customers and recruit new customers in a dynamic and interactive manner is rapidly evolving and having a direct impact on Australia's digital economy and digital citizenship.

The direct contribution of the internet to the Australian economy was worth approximately \$50 billion or 3.6% of Australia's Gross Domestic Product (GDP)

³ Nielsen Australian Online Consumer Report March 2012

⁴ Facebook S-1, 2012

⁵ http://www.youtube.com/t/press_statistics/

⁶ Ibid

⁷ Sensis Yellow Social Media Report May 2013 p53

⁸ Sensis E-Organisation Report 2012, p4

in 2010. This contribution is of similar value to the retail sector or Australia's iron ore exports. Australians are now spending 21.7 hours per week on the Internet.

The not for profit sector is also benefiting from the reach that social media platforms enable, with 92% of not for profits having an online presence and citing LinkedIn, Facebook, YouTube and Twitter as the most frequently used social platforms⁹. Interestingly, not for profits making less than AUD\$100,000 a year in revenue were the most likely to use social media¹⁰.

This new ability for social media users to easily and rapidly coalesce around certain issues, causes, interests and campaigns is fast becoming a hallmark of the way in which Australians come together and express how they feel and what they believe to their friends, colleagues, former employers, charities they support, organisations and government. These virtual conversations are the modern day equivalent of the 'water cooler conversation'.

⁹ Wirth Consulting Research, The State of Social Media Use in Australian Non Profit Organisations H12012

¹⁰ Ibid

Summary of best practice recommendations

(i.e. not legally required)

1. Develop moderation guidelines and publish them on your social media property so that your community is very clear about how behaviour is being managed (see Appendix B for examples of issues to include in your moderation policy). Feel free to build on the platform's community guidelines and acknowledge that the platform operator also has the right to delete user comments which are found to violate the site's terms of service.
2. Consider developing an internal moderation schedule, appropriate to your resourcing levels, which identifies who is moderating which social media properties and at which times. This is particularly useful for organisations sharing moderation responsibilities with external agencies or outsourced moderation businesses.
3. Develop a crisis management plan in the event that an issue arises on your social media platform which needs escalating. This may be as simple as using a flow chart to identify who in your organisation should be alerted about an issue, but could also bring in external stakeholders such as law enforcement agencies, child protection agencies, youth counselling services etc.
4. Moderate the user comments on your branded social media properties to the extent your resources allow. At a minimum, it is good practice for you to review and moderate recently published comments at the same time as posting a new comment.
5. If you don't have the resources within your organisation to moderate user comments, or your internal risk analysis has deemed your use of social media platforms to be high risk, consider hiring a specialist moderation business that have all the necessary clearances and are well versed in conflict management and jurisdictional matters.
6. If you are directly soliciting a response or the creation of user generated content in relation to a provocative or edgy question posted on your social media channels which is likely to elicit controversial responses, ensure you have adequate resources to take extra care to review all responses and any provided user generated content promptly.

7. If your business or product is directed towards children, be aware that there may be specific legal or regulatory requirements that you need to meet, and you should employ moderators who have been through a working with children check or police check and who are trained to identify suspicious behaviour which could be indicative of grooming or other predatory behaviour.
8. Regularly review the tools that are available to you when you develop a presence on social media and consider which tools are appropriate for you to implement (e.g. alcohol brands should always use age gating tools to prevent access by under 18s).
9. Provide feedback to the platform operators around how the tools work and any suggestions for improvement.

Why enter in this conversation?

The IAB recognises that there is some confusion amongst the business community about how to manage user comments on social media platforms and hope that this guidance is useful and instructive of best practice in this area. Based on a careful analysis of existing laws and regulation and industry practice, the IAB believes that user comments directed towards an organisation or social media platform, or to other users who are drawn to a particular organisation, do not constitute advertising. That's not to say that organisations shouldn't pay close attention to the user comments being posted on their social media properties in an effort to identify and potentially remove any offensive, off-topic or illegal posts. The IAB asserts that user comments posted on branded social media properties are not perceived as advertising by users of these properties; however user comments can be converted into a promotional statement through an organisations direct endorsement or expression of agreement. This treatment is consistent with the legal liability standard of 'notice and take down' for hosts and publishers of user generated content, the risk of an organisation becoming responsible for a user comment on its social media properties increases once it has been made aware of it and has an opportunity to review it and take appropriate action promptly.

There is a real risk that organisations who treat user comments as advertising will err on the side of caution and moderate user comments very conservatively, which will adversely impact their presence on social platforms and which arguably undermines the very spirit under which social media thrives.

McKinsey recently estimated that the value of the “social economy” was around \$900B to \$1T¹¹. To ensure that these economic benefits of social media, as well as the many personal and societal benefits are enjoyed by all, it is important that user comments reflect and respect basic community standards which are constantly being tested and are evolving. The IAB and its members recognise that all stakeholders have a role in managing user comments;

- Users should think about the appropriateness of their content before they post it and take responsibility for their comments;
- Platforms should remove comments reported to them which are illegal or violate their terms and conditions and empower organisations using their platforms with tools to assist them in moderating their properties;
- The community should report comments that violate applicable rules; and
- Organisations should engage in responsible moderation of user comments posted to their social media channels (see Appendix B for suggestions on how to develop a moderation policy).

People and organisations are using social media for different purposes

People and organisations are using social media for a variety of reasons. This may include sharing information and content, providing customer support, sharing recommendations and experiences, discussing television programs or sales and marketing. Nobody disputes that organisations should take responsibility for the content that they create and upload onto social media platforms. However organisations have limited control over comments which are not created by them and have not been selected or reviewed by them for their suitability for publishing. Brands will always have to be reactive when it comes to considering the appropriateness of user generated comments. This is the fundamental distinction between traditional advertising mediums, which are not interactive, and social media platforms.

¹¹ http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/the_social_economy

Legal treatment of user comments

For organisations using social media as a communications platform, the Australian Consumer Law prohibits any conduct in the course of trade which is likely to mislead or deceive (section 18) or any representations which are false or misleading (section 29). These prohibitions extend to business use of social media and potentially to misleading or deceptive user comments left on an organisations' social media page where the organisation is aware of such comments and has the ability to remove them and fails to do so. Our uniform defamation laws also extend to publications on the Internet (including dissemination on social media sites) and once an Internet publisher or host is made aware of an allegedly defamatory comment they face real risk if they fail to remove the offending statement or user generated comment.

A common legal standard for platforms managing user generated comments is called 'notice and take down'. This means that once the host or publisher of a comment which is illegal is notified that the comment is there (and how to find it!), the publisher or host is deemed to have actual knowledge that there is illegal content on their service and they have an obligation at that point to investigate the notice and take any appropriate action. This is the standard which is applied under the Broadcasting Services Act 1992 and the Copyright Act.

Law continues to evolve in the area of user generated content and social media, therefore there is a degree of uncertainty in these approaches especially given that many social networks are being operated from other countries and may be applying different legal standards to their platform. Any organisation seeking to rely on these guidelines should seek independent legal advice. Examples of different types of user generated content related escalations range from allegations that a comment is defamatory, contains hate speech, misleading and deceptive, infringes copyright, offensive, abusive, harassing, or threatening.

In recognition of the notice and take down standard, social media sites generally make it very easy for users to report illegal or offensive content for review and have developed very responsive teams working around the clock to respond to any such notices. These reporting tools are typically located alongside each and every user comment and are applied across the entire platform including on all branded pages / channels.

This reactive obligation on the hosts and publishers should also, in our view, be extended to organisations that are creating branded spaces within social media platforms.

The Australian Competition and Consumer Commission (the ACCC) enforces the obligations on organisations not to engage in false and misleading conduct under the Competition and Consumer Act 2010. A pharmaceutical company, Allergy Pathway, was successfully prosecuted in 2011 for failing to act on a previous court order to closely monitor their Facebook and Twitter profiles for customers making false and misleading claims about Allergy Pathway products. Allergy Pathway had been advised by the court to be on the lookout for such claims and to remove any that appeared. Allergy Pathway failed to do so and was therefore found negligent. In his judgment, Justice Finkelstein made the point that "while it cannot be said that Allergy Pathway was responsible for the initial publication of testimonials (the original publisher was the third party who posted the testimonials on Allergy Pathway's Twitter and Facebook pages) it is appropriate to conclude that Allergy Pathway accepted responsibility for the publications when it knew of them and decided not to remove them. Hence it became the publisher of the testimonials¹²." The ACCC has provided some helpful examples at Appendix C to assist organisations in their consideration of issues relating to false and misleading conduct on social media platforms.

The IAB recommends that these good practice recommendations be applied across all social media properties, including blogs and Twitter.

The role of moderation on social media platforms

Many organisations using social media platforms are already keeping a close eye on user comments and may even be investing in internal or external moderators to ensure that any offensive or brand threatening posts are promptly removed. This risk is primarily reputational rather than legal, and organisations have commercial discretion to develop, and the flexibility to evolve, their risk management strategies in this context as some will have fewer resources, access to expertise or a higher threshold for controversy than others. This risk analysis should also consider whether an organisation is using an external social networking platform or an owned, proprietary platform where they are hosting user generated content as there are additional legal risks associated with hosting this content.

Moderation is an activity or process following an agreed policy or set of guidelines to encourage safe and responsible use of an interactive service in accordance with the Terms of Service,

¹² <http://www.accc.gov.au/content/index.phtml/itemId/972417/fromItemId/142>

Acceptable Use Policy or House Rules. Moderation is performed by human moderators or filtering software (or a combination) reviewing content posted by users and removing content or restricting users as necessary, either pre- or post- publication in near real time or following user reports¹³. A table identifying the benefits and drawbacks of each different type of moderation can be found at Appendix A.

We recommend that it is good practice to moderate your presence on social media taking practical considerations into account. Factors to consider include:

- How many user comments does your Facebook page or YouTube channel generate?
- How do you deal with spikes in volume given finite resources?
- How do you manage abusive comments outside of normal organisation hours, for example at night or over the weekend?
- What is an appropriate response to an offensive or inaccurate user post – ignore it, respond to it, delete it, ban the user?
- What about borderline examples where a user is simply expressing an opinion or point of view?
- How can you employ technical tools to assist in managing user comments?
- How is the community policing itself?

Given that organisations may use social media for a variety of reasons with a myriad of factors to consider, they may utilise different approaches to moderation. Be aware that there are increased legal risks associated with moderating user generated content before the content is published. This increased risk arises through the assumption that if you are reviewing everything before it is published then you effectively become aware of any illegal content and are choosing to approve it for publishing. Pre-moderation, as this form of moderation is called, can also disrupt the dynamic flow of interactions and frustrate users.

The IAB believes that the starting point of any deliberation is that user comments are not part of an advertising or marketing communication.

Industry practice suggests that an organisation is not responsible for tracking or moderating user comments that mention or reference them in social media channels that they have not established, unless they have been:

¹³ UK Council for Child Internet Safety Good Practice Guidance for the moderation of interactive services for children 2010 <http://media.education.gov.uk/assets/files/pdf/i/industry%20guidance%20%20%20moderation.pdf>

- a) Specifically requested to track or moderate user comments by a recognised regulatory or enforcement authority;
- b) Responded to or re-tweeted by the organisation (for example, this includes the use of the @ hashtag on Twitter, or the tagging of – or linking to – a brand’s social media property, or presence on other platforms); or,
- c) Later referenced in secondary material (both online and offline) by the organisation for marketing purposes.

Recommendations:

- Develop moderation guidelines and publish them on your social media properties so that your community is very clear about how behaviour is being managed (see Appendix B for examples of issues to include in your moderation policy). Feel free to build on the platform’s community guidelines and acknowledge that the platform operator also has the right to delete user comments which are found to violate the site’s terms of service.
- Consider developing an internal moderation schedule which identifies who is moderating which social media properties and at which times. This is particularly useful for organisations sharing moderation responsibilities with external agencies or outsourced moderation businesses.
- Develop a crisis management plan in the event that an issue arises on your social media platform which needs escalating. This may be as simple as using a flow chart to identify who in your organisation should be alerted about an issue, but could also bring in external stakeholders such as law enforcement agencies, child protection agencies, youth counselling services etc.
- Do moderate the user comments on your branded social media properties to the extent your resources allow. At a minimum, it is good practice for you to review and moderate recently published comments at the same time as posting a new comment.
- If you don’t have the resources within your organisation to moderate user comments, or your internal risk analysis has deemed your use of social media platforms to be high risk, consider hiring a specialist moderation business that have all the necessary clearances and are well versed in conflict management and jurisdictional matters.
- If you are directly soliciting a response or the creation of user generated content in relation to a provocative or edgy question posted on your social media channels which is likely to elicit controversial responses, ensure you have adequate resources to review all responses and any provided user generated content promptly.

- If your business or product is directed towards children be aware that there may be specific legal or regulatory requirements that you need to meet, and you should employ moderators who have been through a working with children check or police check and who are trained to identify suspicious behaviour which could be indicative of grooming or other predatory behaviour.

Tools made available to organisations by social media platforms

Facebook, YouTube and Google+ make administrative tools available to organisations creating branded properties on these platforms.

Please refer to Appendix D for information about Facebook tools, Appendix E for information about YouTube tools and Appendix F for information about Google+ tools.

For platforms that don't offer any additional tools for organisations (e.g. Twitter), if you see a user comment that concerns you report it using the general reporting tools on the site.

Recommendations:

- Regularly review the tools that are available to you when you develop a presence on social media and consider which tools are appropriate for you to implement (e.g. alcohol brands should always use age gating tools to prevent access by under 18s);
- Provide feedback to the platform operators around how the tools work and any suggestions for improvement.

Appendix A – Table outlining different types of moderation¹⁴

| | Description | Benefits | Drawbacks |
|--|--|--|--|
| Professional / human moderators | <p>Staff employed or contracted by the service provider to provide:</p> <p>Pre-moderation: in a pre-moderated service all material/content supplied by users reviewed by the moderator for suitability before it becomes visible to other users;</p> <p>Post-moderation: in a post-moderated service, all material/content supplied by users reviewed after it becomes visible to other users and action taken to remove inappropriate content and warn/ban users who break the rules;</p> <p>Sample moderation: a moderator may ‘patrol’ a number of spaces or otherwise examine a sample of content but not all content is reviewed after publication, and</p> <p>Reactive moderation: in a service of this type moderation will take place only after a report is made.</p> | <p>Can take a very subtle approach, understanding the nuances of what makes a particular kind of content or behaviour inappropriate and how to respond effectively.</p> <p>Pre-moderation can, in theory, prevent disallowed content appearing at all.</p> | <p>Staff are expensive and require proper training in order to moderate effectively. Often impractical for staff to view everything that is uploaded, especially where large volumes of content are concerned.</p> |

¹⁴ UK Council for Child Internet Safety Good Practice Guidance for the moderation of interactive services for children 2010 <http://media.education.gov.uk/assets/files/pdf/i/industry%20guidance%20%20%20moderation.pdf>

| | | | |
|--|--|--|---|
| <p>Automatic scanning</p> | <p>A computer programme scans for words, phrases, properties of pictures and videos and/or patterns of behaviour by users to identify inappropriate content of behaviour.</p> | <p>Can 'look' at huge volumes of content at once, saving time and money.</p> | <p>May lack nuance of a human moderator. For example may flag up harmless pictures of a swimmer because of the presence of bare flesh. Alternatively, may miss inoffensive comments where a usually inoffensive word is used in an offensive context.</p> |
| <p>Community Moderation</p> | <p>User's report or 'flag' content and behaviour which they believe is inappropriate and contravenes the site's terms of use. Reports or 'Flags' are then reviewed by moderators for contravention of the site's terms of use.</p> | <p>Potentially, every user on the site can play a role in enforcing the rules of their online community thereby giving the community a sense of 'ownership' over what happens.</p> | <p>Users may not report inappropriate content before others users including children have had a chance to see it. Some users may not wish to abide by the rules. Different users may interpret the rules inconsistently.</p> |
| <p>Reputation based systems</p> | <p>This is a version of community moderation where the 'reputation' a user has built up on a site (e.g. based on their level of activity, or rating they have been given by other users) gives particular weight to reports they make.</p> | <p>Genuinely empowers users, including children themselves, to become responsible, respected members of an online community with a role in keeping themselves and others safe.</p> | <p>May need to be supported by professional moderators to make sure that users with a high reputation do not adopt a vigilante approach which could lead to some users being bullied.</p> |

Appendix B – Developing a moderation policy

In developing a moderation policy, you should first and foremost consider your audience and what may or may not be considered acceptable conduct to them (e.g. consider cultural differences if you have users in other countries).

You might like to include references to any of the following:

- What kinds of behaviours are acceptable and unacceptable?
- Will you allow the promotion of related products and services on your social media property? You might like to disallow these on the basis that you may be held responsible for another brand's advertising.
- Look out for comments about your organisations' staff and develop a policy for handling those.
- Consider how you want to respond to, and escalate, threats of self harm and potential grooming behaviours. Have resources to refer people to, vetted responses to use and ensure that the right people within your association are involved.
- There is a lot of sensitivity around defamatory user comments and some interesting legal precedents addressing this issue. Be conservative and take any arguably defamatory comments down as soon as you see them.

Appendix C – Australian Competition and Consumer Commission (ACCC) Guidance¹⁵

Social Media

Social media gives large and small businesses a direct way to interact with existing and potential customers, and promote their products and services. Businesses using social media channels like Facebook, Twitter and YouTube have a responsibility to ensure content on their pages is accurate, irrespective of who put it there.

- [Don't make misleading claims on social media](#)
- [Don't allow others to make misleading claims in comments](#)
- [Minimise your risk](#)
- [Monitoring social media pages](#)
- [Responding to false, misleading or deceptive comments instead of removing them](#)
- [ACCC role in enforcement of social media](#)
- [Offer your customers a refund](#)
- [More information](#)

Don't make misleading claims on social media

You must ensure you don't make any [false or misleading claims](#) as part of your marketing and promotional activities. This includes advertisements or statements using any media, including print, radio, television, websites and social media channels like Facebook and Twitter.

There are no specific or different consumer laws or rules in place for social media. Consumer protection laws which prohibit businesses from making false, misleading or deceptive claims about their products or services have been in place for decades. These laws apply to social media in the same way they apply to any other marketing or sales channel.

Examples

1. XYZ Pty Ltd tweets that they are the first Australian company to offer a 100 per cent environmentally friendly car wash service when they have not done any research to support this. It turns out that GHI Pty Ltd has offered the same service for many years. This tweet is likely to be false, misleading or deceptive.

¹⁵ <http://www.accc.gov.au/business/advertising-promoting-your-business/social-media>

2. ABC Pty Ltd pays a celebrity to tweet that she loved staying at one of ABC's resorts. The celebrity has never actually been to this resort. This tweet is likely to be false, misleading or deceptive.

Don't allow others to make misleading claims in comments

You can also be held responsible for posts or public comments made by others on your social media pages which are false or likely to mislead or deceive consumers. In 2011, a court case concluded that a company accepted responsibility for fan posts and testimonials on its social media pages when it knew about them and decided not to remove them.

Examples

1. A fan of XYZ Pty Ltd posts negative and untrue comments about a competitor's product on XYZ's Facebook page. XYZ knows that the comments are incorrect, but decides to leave the comments up on its page. XYZ may be held accountable for these comments even though they were made by someone else.
2. ABC Pty Ltd and DEF Pty Ltd are market leaders in the paint industry. A customer posts on ABC's Facebook page that their paint always lasts much longer than DEF's paint. ABC is unsure if this is true, but decides not to remove the post. It turns out that ABC's paint does not last longer. ABC may be held responsible for this comment.

Minimise your risk

Don't make statements on your Facebook or other social media pages that you wouldn't make in any other type of advertising. If you're unsure about what you can or can't say, seek legal advice.

Monitor your social media pages and remove any posts that may be false, misleading or deceptive as soon as you become aware of them. This is what the ACCC would expect you to do with any other type of advertisement.

Establish clear 'house rules' that apply to the actions of your fans, friends and followers when using your social media pages. These rules should be featured prominently on your social media pages. You should then block users who breach those rules.

Monitoring social media pages

The amount of time you need to spend monitoring your social media pages depends on two key factors: the size of your company and the number of fans or followers you have.

Businesses should keep in mind that social media operates 24 hours a day, seven days a week, and many consumers use social media outside normal business hours and on weekends.

Examples

1. ABC Pty Ltd has 300 staff. As larger companies usually have sufficient resources and sophisticated systems, the ACCC would expect ABC to become aware of false, misleading or deceptive posts on its Facebook page soon after they are posted and to act promptly to remove them.
2. XYZ Pty Ltd has only 10 staff but more than 50,000 Facebook fans. Given the number of people who could be misled by an incorrect post on XYZ's Facebook page, the ACCC would expect XYZ to devote adequate resources to monitoring its Facebook page and to remove any false, misleading or deceptive posts soon after they are posted.
3. DEF Pty Ltd has 12 staff and only 80 Facebook fans. As a small business, DEF is unlikely to have the same resources to dedicate to social media monitoring as a larger company would. Also, given the small number of Facebook fans, there is less potential for widespread public detriment from incorrect posts. Accordingly, the ACCC would not expect DEF to monitor its Facebook page as regularly as the companies in the two previous examples.

Responding to false, misleading or deceptive comments instead of removing them

You can respond to comments instead of removing them, but it is possible that your response may not be sufficient to override the false impression made by the original comments. It may be safer to simply remove the comments.

ACCC role in enforcement of social media

The ACCC can require companies to substantiate any claims on their social media pages, and can take court action where it identifies a breach of the law (or issue an infringement notice in certain circumstances).

While all complaints are carefully considered, the ACCC directs its resources to investigate and resolve matters in accordance with our compliance and enforcement priorities. The ACCC will take a proportionate response in relation to false or misleading comments on businesses' social media sites. We are more likely to pursue cases of false, misleading or deceptive conduct if:

- There is the potential for widespread public detriment if the statement is relied on
- The conduct is particularly blatant
- It is by a business that has come to our attention previously

Appendix D – Facebook tools

Page owners and advertisers must comply with the Facebook advertising guidelines

https://www.facebook.com/ad_guidelines.php

Everyone who uses Facebook must comply with the Facebook community standards

<https://www.facebook.com/communitystandards>

Admin tools

When you are setting up your Page, you should review and set the tools available to Facebook Page admins for proactively moderating content on their Page. These are a moderation blocklist and a profanity blocklist.

Moderation Blocklist

The moderation blocklist allows you to add comma-separated keywords to a blocklist in order to prevent comments that include those keywords from appearing on your Page. When people include blacklisted keywords in a post or comment on your Page, the content will be automatically marked as spam.

Profanity Blocklist

The profanity blocklist also Page administrators to proactively moderate posts. It will block the most commonly reported words and phrases marked as offensive by the broader community.

Content moderation policy

When you are setting up a Facebook Page you may want to develop a comment moderation policy that makes clear to users what types of comments you will allow on your Page and what types of comments you will remove.

Deleting comments & banning users

As a Page administrator you can delete user comments made to your Page and you can also ban users, if they continue to post comments that you delete.

Suicidal posts

You can report suicidal posts to Facebook through

<https://www.facebook.com/help/contact/?id=305410456169423>

More information about these tools can be found in the Facebook Help Centre:

<https://www.facebook.com/help/407894169267325/>

Appendix E – YouTube tools

YouTube community guidelines

http://support.google.com/youtube/bin/request.py?contact_type=abuse

YouTube content ratings

<http://support.google.com/youtube/bin/answer.py?answer=146399>

Details on moderating user comments

<http://support.google.com/youtube/bin/answer.py?hl=en&answer=171666>

How to remove comments from videos:

1. Navigate to the video with the comment you would like to remove
2. Scroll down to the All Comments section and locate the comment that you would like to remove
3. To remove comments that you've made on a video, hover your mouse over your comment. A row of options will appear: thumbs up, thumbs down, reply and a downward facing arrow. Clicking on the arrow will reveal Remove.

You can also change how comments are posted to each of your videos. If you edit the properties for a video, you have the option to allow all comments, require approval for each comment, or block all comments. These steps can also be used to remove comments that you've made on other user's videos.

Filtering comments:

Every day millions of people make choices about what kind of YouTube experience they would like to have, choosing videos, languages, privacy settings, feeds, and more.

"Hide objectionable words" is another option, blocking profanity and racial slurs for those users who prefer not to see them. Users can opt into this by clicking on "Options" next to the Comments header and checking the "Hide objectionable words" box. Users can also choose to hide comments altogether by clicking on "Hide all comments." Your preferences will remain in place until you change them on your browser.

YouTube uses a combination of feedback from users, proprietary technology, and a common sense collection of words to decide what to filter. This list will continue to evolve.

Appendix F – Google+ tools

Google+ Pages additional terms of service

<http://www.google.com/intl/en/+/policy/pagesterm.html>

Google+ user content and conduct policy

http://www.google.com/intl/en_uk/+/policy/content.html

Managing a Google+ Community

<http://support.google.com/plus/bin/answer.py?hl=en&answer=2870379>

Google suggests that people add moderators and invite them to manage content and don't leave communities unmoderated. An owner and a moderator will be able to:

- Add and edit categories
- Remove posts
- Remove members from the community
- Ban members from the community
- Add additional moderators

A moderator will be able to do everything an owner can do to manage the community except delete the community. Steps to add moderators:

1. Desktop: Click members below your community's photo Mobile: Touch the members list displayed below community name
2. Click the drop-down menu to the right of the member you'd like to add as a moderator
3. Click Promote from member to moderator

Through the menu located on the side of each post, you and your moderators can manage your community by:

- Removing posts made by members
- Removing members from the community
- Removed members can join or request to join again.
- Banning members from the community Banned members cannot join or request to join again.

Additional Contributors and Supporters

