

The Legal and Ethical Status of Social Media

Jon Fistein

This chapter considers some of the legal and professional implications of social media use. These are discussed from two perspectives: firstly, how the law is changing to help ensure harmful and false information is not distributed, and secondly how professionals should use social media responsibly. To illustrate each of these, the chapter focuses on proposed reforms to the UK law and the UK General Medical Council's (GMC) guidance for doctors. The principles discussed will apply to other jurisdictions and professional groups, but individual professionals should ensure they are familiar with the frameworks within which they practise.

As will be seen, although social media are relatively new, many of the legal and professional issues they present are not. Social media present new opportunities for the more effective dissemination of health information, offer new ways to communicate with individual patients and the public about health matters, and can provide valuable support networks for vulnerable individuals who might otherwise be isolated. However, they also offer more opportunities for people to be misled by information that is wrong or posted deliberately to cause harm, and they make it easier inadvertently or deliberately to spread information relating to particular individuals. This is not new: misinformation has always been in the public domain, and it has always been possible for recipients of such misinformation to be misled or harmed. Confidentiality and privacy have been safeguarded by laws and professional rules that have been in place for many years. The GMC (and other professional bodies) has guidance in place advising doctors and other professionals to be considered in their communications, to ensure they and their professions remain trusted.

The questions we address here are: do existing laws offer sufficient protection given the ease and scale of misinformation seen on social media, and how should existing professional principles be applied to social media?

The Case for Law Reform

Amy is diagnosed with an eating disorder. She goes online, looking for information about the condition and support from others going through similar experiences. She finds several helpful blogs in which people have documented their own recoveries. She also finds a closed social media group in which people encourage and support each other to maintain dangerously low body weight. Amy is distressed by what she reads and finds it harder to cope with her own eating disorder, losing significant weight as a result. Have the members of the social media group done anything unlawful?

The pace of legal change often lags the pace of technological change. In the UK, new law is either passed by government or developed in the courts. This process can take several years and is often reactive. When new technology emerges, the courts apply the existing rules to the new technology, developing the law where possible to ensure their rulings are consistent with existing law, but also fair in the case in question. However, existing law may not be applicable where there has been a radical change, and periodically there may be the need for central law reform from the UK Parliament to revise and consolidate existing laws to ensure they are fit for purpose. This process is currently under way for social media. We will describe these proposed reforms, focusing particularly on how they aim to prevent social media causing physical and mental harm.

In 2014, the House of Lords considered whether the current laws needed revision given the uptake of social media. Jean-Baptiste Alphonse Karr might have said of its position ‘plus ça change, plus c’est la même chose’. It argued that the then existing laws applied to social media, and ‘there are aspects of the current statute law which might appropriately be adjusted and certain gaps which might be filled [to cover social media]’. It was ‘not persuaded that it [was] necessary to create a new set of offences specifically for acts committed using the social media and other information technology’ [1]. One example, discussed further later on in this chapter, is the offence of encouraging suicide – which is illegal under the 1961 Suicide Act [2] regardless of the medium used for the encouragement.

Its argument was that, at their core, social media share some fundamental similarities with traditional media. In both cases there are content creators, publishing channels, a publishing process, and recipients of the published information. One should therefore expect the legal and professional considerations surrounding social media to be very similar to those surrounding more traditional publication channels. There are existing laws that create civil liabilities and criminal offences for harassment, defamation, threatening communications, and for communications that are grossly offensive, indecent, obscene, or false, which would also apply to social media.

However, all has not ‘remained the same’.

In traditional media there is more control over the publishing process than there is in social medial publication, and this can place safeguards on material that appears in the public domain. For example, in traditional media, authors are commissioned based on their knowledge or expertise in a particular topic, and the content they create undergoes a review process to ensure its accuracy and veracity. Publishing houses place their reputation on the quality of the information they produce and hold themselves to ethical standards that entail not misleading their audiences. Reputable publishers additionally make explicit any editorial or advocacy position they take, enabling reasonably informed readers to be able to distinguish fact from opinion.

The traditional model has its downsides, too: routes to publication can be slow, expensive, and restrictive as only certain ‘approved’ authors may be able to air their views. The net result may be that content is filtered, such that only content deemed worthy of publication may see the light of day. In extreme cases this can lead to oppressive control by publishing oligarchies, limiting individuals’ freedom of expression. Individuals’ access to information may also be restricted, as readers may, for example, be required to pay to enable publishers to recoup their production costs.

Social media have countered these negatives by providing platforms that enable content creators to release material with minimal editorial filtering. This enables readers

to access (and interact with) unfiltered content, during which process they become creators themselves. Access to information is not only more open (and potentially cheaper) than in traditional media, but also is less passive. Readers don't merely seek the information they may need, but information can be tailored to their needs and be 'pushed' to them in real time. Virtual communities can self-organize to offer mutual support and information-sharing. This may seem like a utopian position, where information is free and provided to users when and where they need it, but it can have negative consequences where inaccurate information is presented as fact to readers who may be harmed when they rely on information that is designed to manipulate or mislead them.¹

In the UK, the Law Commission (the statutory body that keeps the law under review and recommends reform where it is needed) recognized 'the difference', observing that:

The revolution in online communications has offered extraordinary new opportunities [for people] to communicate with one another and on an unprecedented scale. However, those opportunities also present increased scope for harm: the physical boundaries of a home now afford no haven to the bullied; the domestic-abuser can exert ever greater control over the life of the abused; many thousands of people can now abuse a single person at once and from anywhere in the world. The examples are many [3].

It published its recommendations in July 2021 [4] concluding, contrary to the earlier House of Lords Committee, that 'the current criminal offences are ill-suited to addressing these harms'. It argued that the expansion of social media, the ease of access to it (as a publisher or a reader), and the ease with which misleading or harmful information can be distributed represented a 'fundamental change' in the way people communicate, which has greatly increased the 'scale and varieties of harms' that may flow from online communications. It argued that existing laws both under-criminalize some behaviours (leaving damaging communications without criminal sanction) and over-criminalize others (permitting prosecutions that may constitute an unjustifiable interference in freedom of expression).

The Law Commission, following an extensive consultation, therefore proposed reform in four areas of law, three of which are the more relevant here:

1. The creation of a 'harms-based' communications offence, to make it illegal to send communications that could cause harm to others.
2. Offences of sending knowingly false, persistent, or threatening communications, making such communications illegal.
3. An offence of encouraging or assisting serious self-harm, similar to the existing offence of encouraging or assisting suicide.²

¹ There has been much public debate about the role of platform providers who host harmful material. Some hosts cast themselves as impartial conduits who are therefore not responsible for any harmful content they host, notwithstanding the fact that they have control over what is posted on their sites. A fuller discussion of the role of platform hosts is outside the scope of this chapter; however, it should be noted that in the UK, the proposed Online Safety Bill, discussed in this chapter, will place a duty of care on organizations hosting user-generated content not to host potentially harmful material, which should provide extra safeguards against the appearance of potentially harmful material online.

² On 4 February 2022 the UK Government Department for Digital, Culture, Media and Sport agreed to accept the Law Commission's recommendations on creating a harms-based

We will explore each of these in turn, considering the aspects of the proposals that are particularly relevant to mental health.

The Creation of a 'Harms-Based' Communications Offence

Recommendation 1

We recommend that section 127(1) of the Communications Act 2003 and the Malicious Communications Act 1988 be replaced with a new offence with the following elements:

- (1) the defendant sent or posted a communication that was likely to cause harm to a likely audience;
- (2) in sending or posting the communication, the defendant intended to cause harm to a likely audience; and
- (3) the defendant sent or posted the communication without reasonable excuse.
- (4) For the purposes of this offence:
 - (a) a communication is a letter, article, or electronic communication;
 - (b) a likely audience is someone who, at the point at which the communication was sent or posted by the defendant, was likely to see, hear, or otherwise encounter it; and
 - (c) harm is psychological harm, amounting to at least serious distress.
- (5) When deciding whether the communication was likely to cause harm to a likely audience, the court must have regard to the context in which the communication was sent or posted, including the characteristics of a likely audience.
- (6) When deciding whether the defendant had a reasonable excuse for sending or posting the communication, the court must have regard to whether the communication was, or was meant as, a contribution to a matter of public interest.

We recommend that "likely" be defined as "a real or substantial risk".'

The Law Commission set out several elements that would need to be proven for an offence to have been committed under the proposed new law. These include:

- defining harm: establishing thresholds for harm and the likelihood of harm;
- the likely audience of any communication;
- the intention behind the communication;
- whether there is a reasonable excuse for the communication; and
- whether there should be a distinction between communications that are 'sent' versus those that are 'posted' on social media platforms.

The practical implications of these regarding harms to mental health are considered later in the chapter.

The Law Commission recommends that 'harm' should be defined as 'at least serious distress', to exclude other, more minor harms to individuals. Although the line between minor and more serious harm may seem nebulous, the concept of 'serious distress' has already been established in English Law, as distress that has 'substantial effect' on the victim's usual day-to-day activities. The Law Commission noted that 'serious distress' would include 'recognised medical conditions' and 'substantial adverse effects' on the

communication offence, a false communications offence, and a threatening communications offence. It is considering whether to adopt the offence for encouraging or assisting self-harm.

victim but should go beyond these as the bar for defining harm in the context of social media should be lower.

The Law Commission recognizes that 'distress is often more difficult to prove than physical harm' and that its broad definitions, if taken in isolation, might incite trivial or vexatious legal actions. To counter this, it argues that the proof of harm would be only one of the several elements required for a successful conviction, all of which would need to be proven, and which only taken together would lead to a conviction. It also recommends that there should be guidance to accompany any law to provide indicative, non-exhaustive examples of the kinds of distress that courts should (and should not) interpret as being within the definition. It notes that harm could follow a single communication, as distinct from the series of communications that would need to have occurred to establish harassment.

It is important to note that for a communication to be an offence under the proposed law, there is no requirement to prove that actual harm resulted but there is the requirement to prove that the communication is 'likely to cause harm' (of the serious nature already described). This is, again, not a new principle, and is seen in other offences such as assault and battery, for which there is also no requirement for actual harm.

This provision is important for two reasons. Firstly, it makes sense to make it unlawful to communicate material that could foreseeably cause harm, as a preventative measure to avert actual harm to potential victims. Secondly, this limits the scope of the offence to exclude cases where it is unlikely that harm would be caused by the communication, even if unexpected harm did occur, thereby limiting the risk of 'vexatious complaint'.

In considering the likeliness of potential harms, the nature of the potential victim needs also to be taken into account. The Law Commission resisted adopting a 'reasonable person' test, which was proposed by some respondents to its consultation who wanted to ensure that causing 'harm' to individuals who are uncommonly and unreasonably sensitive (and likely to be caused distress) was excluded. Instead, it argued that is better to focus on the likelihood of harm resulting from the communication and the potential victim – 'if the sender knows the likely audience is unusually susceptible to a particular type of harm and exploits that to their serious disadvantage, the sender should not be able to argue their innocence by appeals to a reasonable person'. It also argued that this is important to protect potential defendants, who should be able to 'predict at the point of sending whether they are about to commit a criminal offence', that is, whether they are likely to cause 'serious harm' to a likely audience, who may be in a vulnerable group.

As material can be posted on social media without any specific recipient in mind, the Law Commission also recommended that consideration should be given to whether the communication was likely to be 'seen, heard or otherwise encountered' by potential victims. This is to protect potential victims even if they are not specifically targeted, by making it an offence to post harmful material if there is a likelihood that it will be seen by those potential victims (presumably including where they may be 'targeted' by algorithms pushing content to them). The Law Commission recognizes that there needs to be a balance, though, between the need to prevent the posting of information with the intent of causing harm (which would be illegal under the proposals) and the requirements not to disproportionately restrict rights of freedom of expression (as outlined in the Human Rights Act), for example by criminalizing legitimate discussion of contentious or emotive topics, which may lead to harm, even though that is not the intent.

The proposed law therefore considers the intent of the sender as well as the content of the message, proposing that an offence should only be deemed to have been committed if the sender intended to bring about the harm. This excludes recklessness (cases where the sender may be aware of the potential risk of harm but does not intend to cause harm). The Law Commission additionally proposes that for communications to be unlawful, they must be made without 'reasonable excuse': the prosecution would have to establish the lack of such reasonable excuse for the communication for the prosecution to succeed. This should further ensure that there is 'no interference by the criminal law in reasonable communications or legitimate freedom of expression . . . even when someone has sent a likely harmful communication intending to cause harm'. Although this may sound like a high burden and standard of proof, the Law Commission also notes that other legislation is in progress (such as the Online Safety Bill [5]) that will protect potential victims.

Offences of Sending Knowingly False, Persistent, or Threatening Communications

Recommendation 3

'We recommend that the existing offences in sections 127(2)(a) and (b) of the Communications Act 2003 be replaced with a new summary offence targeting knowingly false harmful communications with the following elements:

- (1) the defendant sent or posted a communication that he or she knew to be false;
- (2) in sending the communication, the defendant intended to cause non-trivial psychological or physical harm to a likely audience; and
- (3) the defendant sent the communication without reasonable excuse.
- (4) For the purpose of this offence:
 - (a) a communication is a letter, electronic communication or article (of any description); and
 - (b) a likely audience is someone who, at the point at which the communication was sent by the defendant, was likely to see, hear, or otherwise encounter it.

We recommend the offence should have a limitation period of three years from the day the offence was committed and six months from the day the prosecution becomes aware of sufficient information to justify proceedings.'

As well as the harms-based offence, the Law Commission made several recommendations relating to the posting of false and threatening information. The most relevant of these here is Recommendation 3, shown here, relating to the posting of false information.

The Law Commission was careful to point out that for a communication to be unlawful under this offence, the information must be false and the sender must know it to be false. This would mean that a communication that was false, but which the sender sincerely believed to be true, would not be unlawful. This position was adopted to recognize that the criminal law should only play a role in cases where false communications are made with 'malign purpose': the aim is to ensure the offence targets 'culpable false communications and has robust protections for freedom of expression' [4].

The false communications offence is proposed as a complement to the more general harms-based communication offence already described. It mirrors the harms-based offence in several ways. For example, a key element of the proposed offence is that the

instigator of the communication must have ‘intended to cause non-trivial psychological or physical harm to a likely audience . . . without reasonable excuse’. The relevant legal principles are covered in the previous section.

An Offence of Encouraging or Assisting Serious Self-Harm

Recommendation 14

‘We recommend that encouraging or assisting serious self-harm should be an indictable offence with the following elements:

- (1) A person (A) intentionally encouraged or assisted another person (B) to cause himself or herself serious physical harm.
- (2) B need not be a specific person (or class of persons) known to, or identified by, A.
- (3) A may commit this offence whether or not serious physical harm, or an attempt at causing serious physical harm, occurs.
- (4) Physical harm is serious if it would amount to grievous bodily harm under the Offences against the Person Act 1861.’

‘While encouraging or assisting suicide is [already] a specific offence, criminalised under the Suicide Act 1961, encouraging or assisting self-harm is not’ [4]. The Law Commission proposes a specific new offence to complement both the Suicide Act and its proposed harms-based communication offence, to criminalize the encouragement or assistance of non-suicidal self-injury such as self-cutting, disordered eating, and self-bruising. It draws many parallels to S.2 of the Suicide Act 1961, which sets out the conditions for the offence of the encouragement of suicide and draws on definitions from the harms-based communications offence.

The proposed offence is an ‘inchoate’ one – that is, an offence would still be committed regardless of whether anyone actually self-harmed as a result of the encouragement. Such inchoate offences normally relate to the encouragement of other crimes, but it is noteworthy that the self-harm is not in itself a criminal offence (nor is there an intention to make it one). This is similar to the offence of encouragement of suicide, which remains an offence in its own right, notwithstanding the decriminalization of suicide in the Suicide Act 1961.

The proposed offence targets those encouraging self-harm in others. However, the intent of the communication is important here as well. The sender must have intention of causing harm for the offence to have been committed. This is to ensure the offence does not have a disproportionate effect on vulnerable people by ensuring they are not criminalized or stigmatized if they share information about their experiences with self-harm on social media as part of a support community or when seeking help. ‘Those expressing their own experiences with self-harm without any intention to encourage others to self-harm’ would not be committing an offence.

Similarly to the harms-based communications law, the proposed law on self-harm is limited to cover ‘serious physical harm’. Many of the arguments about the seriousness threshold covered in the section on the harms-based offence apply here too, but there is an additional consideration. There are cases where someone can inflict harm on another, but this is not illegal if the person who is harmed gives their consent. The Law Commission does not desire to criminalize ‘the encouragement of an act that, when actually committed against another, would be legal’. The aim of the proposal is to protect

against ‘at least serious harm . . . [which is] more than transient or trifling harm’, which could lawfully be ‘committed by A against [a] willing [i.e. consenting] B’. Although this approach may seem over-cautious, for example, excluding many cases where the self-harm is ‘disordered eating’, the Law Commission was satisfied that this should be excluded from this proposed offence as it ‘may well be captured by’ the other offences it proposes, as we have already set out. As with those other offences, the Law Commission stresses the need for guidance to accompany any proposed new criminal law, in particular to ensure that ‘acts of vulnerable people are not inadvertently criminalised’ [4].

It is unlikely that the blogs looked at by Amy are unlawful. Even though they are discussing harm, the intent is to promote recovery, not to encourage harm.

Considering the postings of the social media group supporting and encouraging low body weight, under the proposed harms-based law, this may be unlawful. The communication is likely to cause serious distress and/or harm to likely recipients and is likely to reach those recipients as it is posted to a group for people with eating disorders. Whether this is unlawful would most likely hinge on the intent of the posters: if they intended to cause distress or harm (for example, by trying to reinforce a vulnerable person’s already distorted body image by commenting on a picture of them, saying they are fat and giving them strategies to lose more weight) it is likely that this would be found to be unlawful due to their intention to cause the recipient to be harmed (by intending to bring about weight loss that would be harmful, even if they thought they were helping) and it is unlikely that there is a ‘reasonable excuse’ in these circumstances.

Professional Conduct on Social Media

Dr S, a GP, discovers that Dr T is posting information on a social media platform espousing the discredited theory that the MMR vaccination causes autism. She is worried that this misinformation might lead to her patients and the public becoming vaccine hesitant, making their children vulnerable to measles, with potentially serious consequences in terms of their morbidity and mortality. What should she do?

The first section considered what protections should be in place to prevent harmful material being posted on social media. This section looks at social media from the perspective of professionals who may use social media themselves, and who may have to counter misinformation online in discussion with service users. We will take the professional standards expected of doctors, set out by the GMC as our exemplar, but similar guidance exists for other professions such as teachers [6]. It should be noted that professional regulators often expect, for those registered with them, higher standards than the legal standards that apply to the general public.

The GMC has issued specific guidance on the use of social media by doctors [7]. The GMC recognizes many of the benefits of social media, such as enabling public health and policy discussions, establishing professional networks, and facilitating patients’ access to information about health and services. However, it reminds doctors that the ‘standards expected of [them] do not change because they are communicating through social media rather than face to face or through other traditional media, but new challenges can arise’.

We will consider two of these challenges here: how doctors should respond to information that is inaccurate or misleading, and doctors' own conduct on social media.

Responding to Inaccurate or Misleading Information

Doctors may encounter material they believe may cause harm to either to their patients or to the public in general. Doctors have a duty of care to their patients, and part of this is to provide information to them that allows them to make informed decisions about their care. Doctors should therefore be careful to provide reliable sources of information to patients, and to direct them away from sources that are misleading. They have a duty to work with their patients to help them be discriminating about the quality and reliability of the information they rely on, for example by explaining how they can check their sources, consider any potential hidden agendas, and be reassured that the information is not false or likely to cause them harm.

Doctors may have a wider duty to correct misinformation [8]. Where doctors encounter information that they feel is misleading and which is aimed at a more general public audience, they should consider their response carefully. The GMC's *Good Medical Practice* requires doctors to treat colleagues 'fairly and with respect' and requires doctors to ensure their 'conduct justifies [their] patients' trust in [them] and the public's trust in the profession' [9]. Doctors must also maintain appropriate boundaries and should not mix their personal and professional relationships. The GMC reminds doctors that they should not 'bully, harass or make gratuitous, unsubstantiated or unsustainable comments about individuals online' and that the laws relating to defamation (making unjustifiable statements about individuals or organizations that harm their reputation) also apply.

In the face of misleading information doctors should therefore, in most cases, resist the temptation to get directly involved with the poster. Rather, they should assess the impact of the information on their patients, and on the trust of the profession and should work to correct the information with their patients. If the information is posted by another doctor, they should be careful not to get dragged into spats on social media. Concerns could be raised with the GMC, and in the future could potentially come under the proposed 'harms-based' or 'false information' laws we have described.

Professional Conduct on Social Media

Doctors may also use social media themselves. Where doctors post material on social media, the GMC reminds them to consider the implications of what they post: that such information is likely to be taken on trust and may potentially be taken to represent the views of the profession more generally [7]. The British Medical Association has also emphasized that a doctor using social media 'is still a doctor' and that the 'ethical and legal standards expected [of them] by the GMC and the broader, less well-defined professional expectations of their peers can still apply online as in any other part of everyday life' [10].

Doctors should also consider the implications of their social media use in relation to laws on privacy and confidentiality [11]. Both of these laws apply to social media, and the ease by which information can be transmitted, copied, disseminated, etc. can increase the chance and extent of potential breaches.

Privacy is the (qualified) right for individuals to conduct various aspects of their private and family life without unwarranted interference [12]. Doctors should consider their own privacy and also that of their patients. As online information can potentially be accessed by others, doctors should be careful to consider whether information they post might lead to breaches of privacy for themselves or others. They should particularly be aware that there might be additional risks to privacy due to the nature of the medium, such as the disclosure of location information in metadata that is associated with posted images and should use appropriate safeguards to protect against these.

Confidentiality is an important professional and legal duty for doctors, and its principles apply when using social media in other settings. It is obvious that doctors should take care to ensure that they do not use publicly accessible social media to discuss individual patients' care, even with those patients themselves, due to the potential risk that the conversation might be intercepted or otherwise compromised. Doctors should also be aware of the risks of 'jigsaw' identification [13], where 'individual pieces of information [that] do not breach confidentiality on their own' are shared but, when taken with other information published online, could be enough to identify a patient or someone close to them.

In summary, social media provide potentially useful ways for doctors, patients, and the public to communicate with each other, and existing professional guidance recognizes this. However, doctors and other professionals must take appropriate care to ensure they are aware of the potential risks and that appropriate safeguards are in place.

Dr S should:

- talk to her patients, to explain the evidence for the safety of the MMR vaccine and its efficacy in preventing serious illness;
- report the misleading information to the social media provider;
- report Dr T's post to the GMC; and
- consider posting accurate information to counter the misinformation.

Dr S should not:

- post anything that personally attacks, disrespects, or defames Dr T.

Conclusion

- Misinformation has always been out there. It is more easily spread now through social media as it is unfiltered and can appear to target vulnerable individuals.
- Many of the existing UK laws provide some protections against the posting of harmful and misleading information.
- However, it has been recognized that these leave some loopholes, so there are proposals for UK law reform to:
 - o place responsibilities on social media platform providers to not allow the posting of harmful or misleading information; and
 - o introduce new offences of:
 - encouraging harms;
 - false communications; and
 - encouraging self-harm – potentially.

- When faced with misinformation online, these may provide routes for professionals to report crime to the police, and further action may be taken by them to investigate and then recommend prosecution.
- Professionals should take care when considering whether to intervene directly if they discover misinformation. They should consider:
 - their professional obligations relating to their behaviour and the reputation of the profession;
 - the impact of the misleading information on their patients, and how they can work with their patients directly to counter any misinformation; and
 - the impact of any action they take on their patients' trust in them and the public's trust in their profession.
- Professionals should also consider their own behaviour on social media, considering, for example:
 - privacy and confidentiality, particularly when information is being stored on remote systems over which they do not have control; and
 - their professional obligations relating to their behaviour and the reputation of the profession.

Further Reading

British Medical Association, *Social media, ethics and professionalism*, www.bma.org.uk/media/1851/bma-ethics-guidance-on-social-media-2018.pdf

Law Commission, *Modernising Communications Offences* (Law Com No. 399, 2021).