The legal implications and challenges of deepfakes

Published 4 September 2020

Intellectual Property analysis: Kelsey Farish, solicitor specialising in technology at DAC Beachcroft, considers the issue of 'deepfakes', its legal implications and the challenges it poses to intellectual property rights. She also considers how UK law is set up to deal with deepfakes and how one might limit their exposure to the risk of deepfakes.

This article was originally published on LexisNexis.

What are the legal implications of audiovisual manipulation and deepfakes and what challenges do they pose to intellectual property rights, rights in personal information and image rights?

The term 'deepfake' describes a face-swapping technique, whereby images of an individual are used by artificial intelligence (AI) technology to generate digital doppelgängers (look-alikes) and then superimposed onto different bodies. Deepfakes generated with only one source image are often obvious as fakes, but those generated with thousands of images or video clips can be very realistic. In contrast to deepfakes, other forms of audiovisual manipulation which do not utilise AI are known as 'shallow fakes' or 'cheap fakes'.

Deepfakes pose many potential systemic socio-political risks such as manipulation of civil discourse, interference with elections and national security, as well as the erosion of trust in journalism and public institutions more generally. The harm posed to individuals and companies are no less important to consider, and include false endorsements, fraudulent submissions of documentary evidence, loss of creative control over audiovisual content, extortion, harassment and reputational damage.

How can ordinary people as well as celebrities limit their exposure to the risk of deepfakes?

There are two broad categories of deepfake risk: the first concerns being deepfaked (ie having your image used in a deepfake video), whereas the second concerns being misled into believing that a deepfake is genuine.

With respect to the first, it is important to remember that deepfakes use images and videos of an individual as the source material to generate digital doubles. Celebrities will have a wealth of material capable of being used for this purpose, but non-celebrities too will likely have a prominent online persona comprised of images. To minimise this risk, it would be prudent to post selfies and other portraits only to private accounts, or otherwise limit the quality and quantity of any such images shared publicly. However, this may not be appropriate in all situations, especially for public figures, influencers, and others who are keen to promote their personal brand.

Images which show an individual at different ages, to include those used for flashback or 'on this day ten years ago' memes provide AI algorithms valuable information regarding how people age and should be shared with caution. The public sharing of images depicting children, even one's own, is generally not advised. Exceptional care should be taken with respect to using mobile apps for face-swapping or augmentation purposes: these apps are often based outside of the UK and EU, and their compliance with data protection legislation is notoriously difficult to enforce.

As for the second key risk, solutions to detect deepfakes are subject to ongoing development and improvement. Some mitigation efforts include meta-tagging source images, watermarking genuine videos (to distinguish them from fakes), or requiring two-factor authentication prior to making decisions which may be fraudulently requested by a deepfaked video or audio clip. Until widespread and reliable solutions are in place however, emphasis on digital literacy to include training for employees in high-risk occupations may be the most practical approach.

How is UK law currently set up to deal with deepfakes?

Legislation specific to deepfakes does not exist as such in the UK. However, other more established laws and legal doctrines may be applicable when attempting to resolve disputes concerning an unwanted deepfake or manipulated video.

Although image rights are not formally recognised in the UK, English case law has in some instances recognised certain protections against the commercial misappropriation of one's publicity. The most relevant of these is likely the tort of passing off, although this may be unworkable for a victim who has not previously commercialised their image. Success turns on specific facts and will depend on whether the individual has a significant reputation or goodwill, and whether the deepfake falsely suggested, to a significant section of the relevant market, that they made a commercial endorsement (see

Irvine v Talksport [2003] EWCA Civ 423 and *Fenty v Arcadia Group* [2013] All ER (D) 410 (Jul) and [2015] All ER (D) 157 (Jan)).

Deepfakes may give rise to a claim for malicious falsehood, but only insofar as they contain false words which result in quantifiable monetary loss. Published images which damage an identifiable individual's reputation may constitute defamation, but following reforms under the Defamation Act 2013, such a publication must normally cause 'serious harm' to the individual depicted in order to be actionable.

Privacy laws including the Data Protection Act 2018 or the General Data Protection Regulation, Regulation (EU) 2016/679 are unlikely to apply in a strict sense, as deepfakes are most often generated using photographs made available by the victim (eg through sharing selfies on social media). Furthermore, although they may be realistic, deepfakes inherently depict a hypothetical fantasy scenario for which privacy will not be at issue as such. Nevertheless, where deepfakes contain confidential details (for example, if a deepfaked person 'speaks' to reveal private facts about another) the torts concerning breach of confidence and misuse of private information may apply.

Copyright protections may likewise be difficult to assert, given the multitude of potential rights holders concerned and potential for fair dealing exemptions. It is also important to note that complications will likely arise as these rights and protections must be balanced against freedoms of expression.

As for image-based sexual abuse (the preferred terminology which includes harms such as 'revenge porn'), the Law Commission is now conducting a review of the existing criminal law with respect to taking, making and sharing intimate images without consent. This specifically includes potential revisions to section 33 of the Criminal Justice and Courts Act 2015, voyeurism offences under section 67 of the Sexual Offences Act 2003 (SOA 2003), the Voyeurism (Offences) Act 2019, exposure under SOA 2003, s 66, as well as the common law offence of outraging public decency.

How are social media platforms taking action to combat deepfake creation and proliferation and how can they stop the spread across jurisdictions of a problematic deepfake?

As of August 2020, several popular social media companies and websites have officially banned or plan to ban deepfakes from their platforms. These include Facebook, Instagram, Twitter, PornHub, and Reddit. Tiktok once embraced the technology, but following discussions concerning the Chinese company's potential take-over by US corporation, has since stated that deepfakes are now banned. That said, Snapchat spent US$ 160m on the acquisition of deepfake technology company AI Factory in 2019.

Regardless of whether or not deepfakes are officially banned, removal or moderation of deepfakes carries with it several complications. Many platforms have been slow or in some instances unwilling to act as arbiters, and some of the bans have wide loopholes, meaning that the prohibition on deepfakes is often in name only. Firstly, from a practical perspective it is often time and labour intensive to identity a deepfake or its creator. There is also no bright line to separate deepfakes from videos with 'acceptable' forms of augmentation or special effects. Secondly, removal of such content may have a chilling effect on free speech. Thirdly, studies show that the spread of novel and entertaining content, even where false, is more likely to be shared. It may therefore be within a platform's commercial interests to keep questionable content online. Fourthly, unlike explicit hate speech for example, the extent to which a deepfake is 'harmful' as opposed to acceptable will largely be down to subjective context, and possibly the opinion of the person(s) depicted. Finally, given the cross-border nature of the internet and the lack of harmonised laws on image rights and media, it is exceedingly difficult to stop the spread of deepfakes across jurisdictions.

Interviewed by Sabina Habib.

Authors

Kelsey Farish
London - Walbrook
+44 (0) 20 7894 6320
kfarish@dacbeachcroft.com