European and UK Deepfake Regulation Proposals Are Surprisingly Limited

Originally published at https://www.unite.ai/european-and-uk-deepfake-regulation-proposals-are-surprisingly-limited/

Analysis For campaigners hoping that 2022 could be the year that deepfaked imagery falls within a stricter legal purview, the early indicators are unpromising.

Last Thursday the European Parliament ratified amendments to the Digital Services Act (DSA, due to take effect in 2023), in regards to the dissemination of deepfakes. The modifications address deepfakes across two sections, each directly related to online advertising: amendment 1709 pertaining to Article 30, and a related amendment to article 63.

The first proposes an entirely new article *30a*, titled *Deep fakes*, which reads:

> *'Where a very large online platform becomes aware that a piece of content is a generated or manipulated image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful (deep fakes), the provider shall label the content in a way that informs that the content is inauthentic and that is clearly visible for the recipient of the services.'*

The second adds text to the existing article 63, which is itself mainly concerned with increasing the transparency of large advertising platforms. The pertinent text reads:

> *'In addition, very large online platforms should label any known deep fake videos, audio or other files.'*

Effectively, the legislation seems to be preparing for the growing practice of 'legitimate deepfakes', where permission has been granted and rights secured for face-swapping in promotional or advertising material – such as Russian telco Telefon's licensed use of Bruce Willis's identity in a recent advertising campaign.

Hesitation to Legislate

The DSA, thus far, does not seem to address the concerns of campaigners over the use of deepfake techniques as they are most commonly used – to re-cast pornographic videos. Neither does it address the extent, if any, to which the pending use of deepfakes in movies and television will need to be disclaimed to viewers in the same way as deepfakes in advertising will be, at least in the EU, from 2023.

The ratification process for the DSA now passes on to negotiation with EU member states, together with the broader scope of the Digital Markets Act (DMA).

Europol's December 2020 report *Malicious Uses and Abuses of Artificial Intelligence* asserted that it would be a mistake for the EU to address specific current deepfake technologies (such as DeepFaceLive), which might lead to EU law constantly playing catch-up with the latest framework or method.

The report stated:

> *'In particular, those policies should be technology-agnostic in order to be effective in the long run and to avoid having to review and replace these on a regular basis as the technology behind the creation and abuse of deepfakes evolves.*
>
> *'Nevertheless, such measures should also avoid obstructing the positive applications of GANs.'*

The concluding remark in the quote above, regarding Generative Adversarial Networks (GANs) broadly characterizes European and North American hesitation to apply laws that might hamstring an emerging AI research sector already perceived to be falling behind Asia (whose more didactic nations have been able to fast-track deepfake legislation).

For instance, a 2018 report from the UK's Select Committee on Artificial Intelligence at the House Of Lords emphasizes several times the risk of allowing timidity to hold back AI development in the nation, exemplified in its title: *AI in the UK: ready, willing and able?*. Last April, Britain also became the first country to green-light the deployment of self-driving cars on motorways.

America is no less avid; in the US, The Brookings Institution has urged the need for increased legislation for AI in the United States, lambasting lawmakers for their 'wait and see' standpoint on the ramifications of machine learning technologies

Besides the insipid approach of the DSA toward addressing social (rather than political) concerns around deepfakes, the EU's proposed regulatory framework for AI, released in April 2021, came under prompt criticism for its own evasion of the topic.

Scant Deepfake Regulation in the UK

As an additional disappointment for anti-deepfake campaigners such as author Helen Mort, who campaigned prominently for new UK legislation in 2021 after being non-consensually depicted in pornographic deepfake videos, a report published today by the UK Parliament's Digital, Culture, Media

and Sport Committee criticizes the British government for failing to address deepfakes in the Draft Online Safety Bill.

Citing the draft bill's current legal redress against deepfake abuse as 'unclear and impractical', the report suggests that the proposed legislation does nothing to address the 'legal but harmful' status of AI-assisted pornographic video and image manipulation techniques:

> *'[We] recommend that the Government proactively address types of content that are technically legal, such as insidious parts of child abuse sequences like breadcrumbing and types of online violence against and women and girls like tech-enabled 'nudifying' of women and deepfake pornography, by bringing them into scope either through primary legislation or as types of harmful content covered by the duties of care.'*

Current applicable law in the UK is confined to the dissemination of 'real' images, such as cases of revenge porn, where, for instance, confidential and private explicit material is publicly shared by an ex-partner. If a persecutor carries out and publicizes deepfake material that superimposes their 'target's' identity into pornographic content, they can only be prosecuted either if they directly harass the target by directing the material at them, or under copyright-related legislation.

In the first case, the ease with which new deepfake content gathers traction and viewers almost inevitably means that the victim will be informed by concerned friends or unrelated third parties, rather than by the person who deepfaked them, allowing the virality of such material to protect the deepfaker, whose work still 'reaches the target'.

In the latter case, prosecution would only likely be feasible where an undoctored third-party pornographic video (into which the victim's identity is later superimposed) is professionally produced and legitimately protected under UK copyright domain (even though a suitable video may be sourced freely from any legal jurisdiction in the world). An 'amateur' video from any jurisdiction lacks clear copyright status, and a bespoke video that the deepfaker has shot expressly to superimpose the victim into is (ironically) itself protected under copyright laws, so long as it complies with other laws.

Behind the Curve

In December of 2021 the UK's Law Commission proposed to extend hate speech laws to cover sex-based hostility, but did not propose the inclusion of deepfakes in this category, despite multiple examples of such usage around the world (especially in India) of the technology being weaponized against female politicians and women activists. Women are overwhelmingly the target of illicit deepfake content, whether the motives of the fakers are overtly social (i.e. the intention to humiliate, de-platform, and disempower) or simply prurient (i.e. pornographic) in nature.

In March of 2021 the Illinois-based National Law Review took the UK's legal framework to task as 'wholly inadequate at present to deal with deepfakes', and even lacking in basic legal mechanisms that protect a person's likeness.

Deepfake Laws in the United States

By contrast, the United States does to some extent protect its citizens' 'Right of Publicity', though not at a federal level (at present, such statutes exist in approximately half of US states, with wildly varying legal mechanisms).

Though an improvement on the UK's performance in deepfake legislation, the US can only boast sporadic, per-state coverage, and seems determined to address the technology's potential for political manipulation before getting round, eventually, to its impact on private individuals.

In 2019 the State of Texas outlawed the creation or spreading of *political* deepfakes, with Texas Senate Bill 751 (SB751), omitting any statement about deepfake pornography. The same year, the State of Virginia added an amendment to an existing law regarding the *Unlawful dissemination or sale of images of another*, appending the broadly encompassing term *'falsely created videographic or still image'*.

In 2020 the State of California enacted *California Assembly Bill 602* (AB 602) prohibiting the generation or dissemination of pornographic deepfakes. The Bill has no sunset clause, but has a statute of limitations of three years, and is accompanied by a separate clause covering political deepfakes.

At the end of 2020 the State of New York passed senate Bill S5959D, which not only outlaws the creation and/or republishing of pornographic deepfakes, but actively protects a user's right of publicity in regard to a *computer-generated* likeness through deepfakes, CGI, or any other means, even after death (if the person in question was a resident of New York at the time of their death).

Finally, the State of Maryland has amended its laws around child pornography to encompass and criminalize the use of deepfakes, though not addressing the impact of deepfakes on adult targets.

Waiting for 'DeepfakeGate'

History indicates that the damage that new technologies may engender has to become *personal* to a nation in order to speed up its legislative response. The very recent death of a teenage girl in Egypt who was allegedly being blackmailed with deepfake pornography of herself has received limited coverage in western media*, while revelations about the theft of $35 million dollars in the United Arab Emirate, which came to light in 2021, also represent a 'distant event' that's not likely to speed up the senate, or light a fire under the 45 remaining states that have not yet enacted deepfake legislation.

If the US adopts a more united front around the abuse of deepfake technology, widespread legislation would likely affect the governance aspect of telecommunications and data infrastructure and storage, leading to rapid catch-up policy changes imposed on its business partners around the world. The fact that Europe's adoption of GDPR did not ultimately 'cross over' to North American data-gathering and retention policy doesn't mean that the EU could not likewise gain leverage over the less compliant nations that it trades with – should it ever take a more committed legislative stand on the generation, storage and retention of deepfake pornography.

But something has to happen on 'ground zero' first, in one of these leading groups of countries; and we are still waiting for it: a colossal darknet haul of CSAM by the authorities; a major heist using audio and/or video-based deepfake technologies to dupe an *American* company director into misdirecting a very large amount of money; or an American equivalent of the growing use of deepfakes to victimize women in more patriarchal countries (if, indeed, US culture is really outfitted to mirror these events, which is questionable). These are hard things to wish for, and good things to avoid by *any other method* than sticking one's head in the sand, or waiting for an 'incendiary' event.

One central problem, which the EU is currently skirting by directing its legislative prowess at advertising companies that *want* to promote their clever and legitimate deepfakes, is that deepfakes remain difficult to spot algorithmically; most of the slew of detection methods that surface at arXiv every month depend on watermarking, blockchain-based verification, or in some way changing the entire infrastructure that we

currently use to freely consume video – solutions which imply a radical legal revision of the notion of video as a proxy for ‘truth’. The rest are routinely outpaced by ongoing advances in the popular open source deepfake repositories.

A further problem is that the major western nations are quite right, in one sense, to not offer a knee-jerk reaction to a single problematic strand in a raft of new AI technologies, many of which promise immense benefit to society and industry, and many of which could be adversely affected in some way if hot-headed proscription and regulation of image synthesis systems were to begin in earnest, in response to a major event, and to the ensuing outcry.

However, it might be a good idea to at least speed up the slow and sometimes aimless walk we are taking towards the regulation of deepfakes, and meet the potential problems in the mid-ground, and on our own terms, instead of being forced by later events into a less considered response.

** The alleged perpetrators are being charged with blackmail; there is no Egyptian law covering deepfake pornography.*

First published 24th January 2022.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai

Discover More