Deepfake Pornography: Beyond Defamation Law
Aasha Shaik

Deepfakes are increasingly present across the internet, with Sensity AI finding that the number of fake videos online has roughly doubled every six months since 2018. While there has been discussion of the danger of deepfake technology to politics, disinformation, and democracy, deepfakes are also a critical matter of women’s rights and gender based violence. In 2019, AI firm Deeptrace found that 96% of deepfake videos were pornographic — nearly all of which manipulated images of women. While deepfakes began as editing celebrities into pornography, the technology is increasingly accessible, and therefore has also become a means of “revenge porn,” also known as image-based sexual abuse, for so-called regular people. While women are often unjustly blamed for revenge pornography — “we should say to...kids in kindergarten really, be careful when transmitting photos,” as Nancy Pelosi put it — deepfakes pose a new threat. Not only can people spread sexually explicit photos of women in ways unintended by the woman, but now, people can do the same using completely PG photos from anyone’s social media.

Sexually explicit media online can cause women severe repercussions, including barriers to or loss of employment, harassment, social isolation, and threats or acts of violence — not to mention the inherent mental toll of trauma. Image-based sexual abuse is an evidently severe issue from which people deserve legal protections. Given the lack of consent involved in the sharing of the content and its sexual nature, many argue that revenge pornography is a sexual offense and should be treated as such legally. The same should apply to nonconsensual deepfake pornography: regardless of whether the media is real (as with traditional image-based sexual abuse) or not, it is a matter of a lack of consent in a sexual context. While some argue that defamation law can be an adequate avenue for pursuing cases of deepfake pornography, defamation laws are insufficient both because they apply to too narrow a subset of image-based sexual abuse, and because they ignore the core issue of consent that is at stake with this form of abuse.

Pursuing a defamation case against non-consensual deepfake pornography would mean arguing that pornographic content is damaging to women’s reputations. But such an argument perpetuates the patriarchal notion that it is wrong for women to be expressly sexual. Furthermore, it ignores the core issue that is at stake with deepfake pornography, which is not whether or not it is “false” or defamatory in its depiction. The central issue is not the damage to a women’s image by being depicted sexually; it is the violation of consent. Obviously, in our current society, women’s reputations, careers, and lives are far too often damaged by expressions of sexuality — but even in an alternate universe where it would cause a woman no measurable harm, socially or professionally, for there to be a video of her engaging in sexual activity online, the production, sharing, and hosting of such a video without permission would be wrong as a violation of consent. In another situation, if someone makes a deepfake or otherwise posts a nonconsensual video of someone who does sex
work, the defense could argue that because the woman’s reputation already involves sexual activity, her reputation is not being defamed by another video of sexual activity. Such an argument — and thus pursuit under defamation as a whole — would be missing the actual point, which is the violation of consent.

Because of defamation laws’ failure to address the core harm, these laws also end up being inadequate from a technical perspective. To prove defamation in a U.S. court of law, the statement (or video) must purport to be true. It would be trivially easy for producers and distributors of non-consensual deepfake pornography to skirt this issue entirely by simply posting “fake” in the title, without ever addressing the core problem posed by deepfakes — that is, the lack of consent. Similarly, legal pursuit under defamation protections could in practice require the deepfake to be a certain degree of realistic — i.e., the defense could argue that because a particular video is shoddily edited and clearly fake, it cannot constitute defamation since the falsehood is evident/not believable. However, how realistic the video is should be immaterial: regardless of whether it can or cannot be easily discernible as fake and therefore does or does not constitute believable harm to a reputation, the deepfake is facilitating a viewing of a person sexually without their consent. Looking beyond deepfakes specifically, the element of falsehood makes defamation a non-comprehensive legal path for various cases of image-based sexual abuse. For example, the aspect of falsity means defamation offers no recourse to real videos of sexual activity posted as revenge.

And even if we do stick to deepfake pornography alone, defamation involves not only falsehood, but distribution of the falsehood. As such, defamation law would limit who can be held liable for deepfake pornography. As with much of online content, there are multiple steps and parties involved in deepfake pornography being online: the creation of the video, its uploading, and its hosting. I would argue that the parties at each of those steps are culpable for violations of consent. But defamation law would only hold culpable the distributor of the deepfake, which not only lets those who played some of those roles fall through the cracks, but fails to protect a woman’s right to not have deepfakes of her privately produced and consumed. Even if a man creates a pornographic deepfake of a woman and does not post it anywhere online or even show it to anyone else, he is still violating the woman’s consent by manipulating her image sexually without permission to do so. That may beg the question of to what extent people retain the right to give or deny consent for the way publicly posted pictures are used. While there is certainly a balance, if we give up our power of consent solely by virtue of posting a photo publicly, we are forcing women to choose between having any sort of public presence and having our faces sexually manipulated.

Another reason defamation is an ill-suited pursuit for cases of deepfake pornography is the lack of automatic anonymity, as compared to sexual offenses. In the UK, for example, victims of other sexual offenses
have automatic anonymity but those of image-based sexual abuse do not — limiting the number of people who report deepfake pornography due to valid fears of harassment, retaliation, stigma, and a loss of privacy.

Deepfakes are yet another example of technology growing exponentially faster than our laws, leaving people already at greater risk of harm without legal protection. While some have assured that we can, in fact, find legal protection under defamation, using defamation protections for cases of non-consensual deepfake pornography perpetuates harmful underlying gender stereotypes, without providing adequate protection against image-based sexual abuse. Indeed, as we have seen, even the narrow issue of deepfake pornography alone is not addressed in any adequate way by defamation laws. The more prudent legal path would therefore be to address the core of the issue — a lack of consent — and treat non-consensual deepfake pornography as what it is: a sexual offense.