

Deep Fake : Régulation française et européenne

Ces derniers mois, le grand public était sensibilisé au réalisme de la technologie du deepfake par des images ou vidéos humoristiques telles que des montages présentant le pape en doudoune ou des chefs politiques dans des situations ubuesques.

Loin de se limiter à des usages humoristiques, les potentielles utilisations malveillantes de cet outil sont nombreuses. Un deepfake visuel et auditif a conduit un employé d'une multinationale hongkongaise à transférer 25 millions de dollars à des cybercriminels se faisant passer pour des membres de la société au cours d'une visioconférence¹. L'usurpation de la voix et de l'image a été rendue possible par la simple utilisation des vidéos de conférence de l'entreprise disponibles sur Internet pour alimenter l'IA. Par ailleurs, il faut s'attendre à une augmentation de l'utilisation de l'IA à des fins préjudiciables tant les progrès de ces outils sont exponentiels².

Cette question a fait l'objet d'une actualité légale récente. En effet, les législateurs français et européens se sont emparés du sujet par la loi du n°2024-449 du 21 mai 2024 et l'AI Act. Ces deux textes, bien que différents dans leur approche semblent se compléter utilement.

I. L'appréhension répressive du droit français

La loi n°2024-449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique est venue intégrer la question de l'intelligence artificielle dans le Code pénal aux articles 228-8 et 228-8-1 du Code pénal.

L'article 228-8 du Code pénal réprime l'usage de deepfakes comme outil de désinformation. À titre liminaire, il peut être relevé que le champ d'application matériel de ce délit ne se limite pas à la publication de contenus en ligne mais également à la diffusion de deepfakes entre particuliers. Dès lors, bien que son appréciation soit large, la caractérisation

¹ Solutions Numériques. « Deepfake, un centre financier chinois se fait arnaquer 26 millions de dollars » (En ligne). Disponible sur : <https://www.solutions-numeriques.com/deepfake-un-centre-financier-chinois-se-fait-arnaquer-26-millions-de-dollars/>. Consulté le : 13 juillet 2024.

² Touraj Ebrahim. « . Intelligence digitale : lorsque le Fake prend le dessus sur le Deep. » (En ligne) Disponible sur : <https://avisdexperts.ch/videos/view/15892>. Consulté le : 13 juillet 2024

d'une « communication de contenu » est impérative pour retenir cette infraction. Ainsi il faut en déduire que la génération de contenus par deepfake, quel qu'ils soient, n'est pas en soi répréhensible. Pour faciliter la poursuite de cette infraction, le législateur incrimine la transmission du contenu litigieux par « *quelque voie que ce soit* ». Toutefois, cette communication peut laisser peu ou pas de traces, notamment entre particuliers.

Afin de garantir l'effectivité de ce délit, le législateur aurait pu opter pour interdire la génération de contenu généré algorithmiquement sans consentement tout en admettant un certain nombre d'exceptions tels que les usages personnels humoristiques visant des personnalités ayant une notoriété publique pour lesquels la nature algorithmique apparaît de façon plus « évidente ».

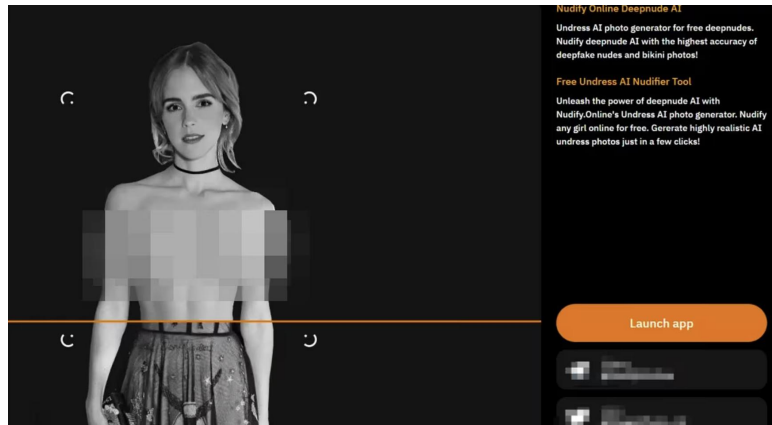
Le législateur a fait le choix de conditionner la caractérisation de ce délit au standard juridique de l'évidence, empreint d'une forte subjectivité. De toute *évidence*, l'évidence de la victime ne sera jamais celle du juge, d'un particulier ou d'un public en ligne. Cette notion doit être utilement entendue dans un contexte éclairant la nature du contenu. Un utilisateur en ligne d'un réseau de partage de contenus de deepfakes n'aura pas la même perception que celle d'un utilisateur moyen de réseaux sociaux grands publics

La question peut être plus délicate si la publication a été initialement diffusée au sein d'un public restreint spécialisé dans l'IA puis communiquée par un de ces membres dans un nouveau public néophyte. À l'*évidence*, il semble que seule la seconde personne puisse être condamnée. Toutefois, le premier auteur a bien « porté à la connaissance du public » le contenu, la nature de ce public n'étant pas explicitement précisée. Ce sont des questions pour lesquelles il est probable que la jurisprudence apporte des réponses dans le futur.

Le traitement des deepfakes sexuelles - La question des deepfakes utilisées à des fins pornographiques est en réalité l'enjeu principal de la réglementation pénale de cette technologie. En effet, selon l'association Deeptrace, 96% des vidéos deepfake sont des vidéos pornographiques dont 99% représentent des femmes.³ Afin d'apporter une réponse à cet enjeu de lutte contre le sexisme, le législateur a consacré à l'article 226-8-1 du Code pénal un

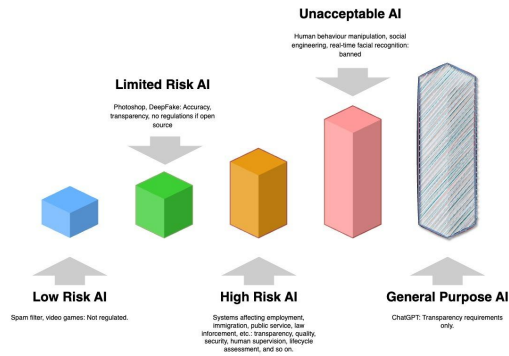
³ Sénat. (En ligne). Disponible sur : https://www.senat.fr/amendements/2022-2023/778/Amdt_128.html. Consulté le 13 juillet 2024.

nouveau délit. Cette infraction interdit toutes les deepfake communiquées au public ou à un tiers dès lors qu'elles ont été réalisées sans consentement et sont de natures pornographiques. Cette sévérité est la bienvenue tant le sujet est important. Toutefois, il peut être regretté de ne pas sanctionner la production sans consentement de ce contenu indépendamment d'une connaissance du public ou à un tiers.



France Info, “"Une seule photo suffit" : quand l'intelligence artificielle démultiplie les deepfakes pornographiques”, Capture d'écran d'un logiciel grand public qui a recours à l'IA pour générer des montages photos. (https://www.francetvinfo.fr/replay-radio/le-choix-franceinfo/reportage-une-seule-photo-suffit-comment-lutter-contre-les-deepfakes-pornographiques-qui-se-multiplient-sur-les-reseaux-sociaux_6213972.html)

II. L'appréhension préventive du droit européen



LAWANDTECH, AI Act is Finally Approved: An Ultimate Regulation, or Will We Keep Needing More Revisions? (<https://lawandtech.ie/ai-act-is-finally-approved-an-ultimate-regulation-or-will-we-keep-needing-more-revisions/>)

Le législateur européen a fait le choix de réglementer l’IA par une approche “par les risques” en distinguant quatre types d’IA classifiés selon leur niveau de risque qu’elle présente allant “d’inacceptable” à minimal”.

L’AI Act consacre à l’article 52 au sein de son paragraphe 3 une obligation de transparence à la charge des utilisateurs d’un système d’IA obligeant ceux-ci à préciser que le contenu à été généré ou manipulé artificiellement.

Par « fournisseurs », l’AI Act désigne les professionnels. L’union européenne opère ainsi une délégation à ces acteurs de la responsabilité de cette réglementation. Or, la création de deepfakes n’est pas l’apanage de professionnels disposant de moyens financiers importants. Cette technologie est accessible aux particuliers. Dès lors, les obligations de transparence instituées par ce règlement ne s’appliquent plus dans les cas n’impliquant pas de professionnels. Cette exigence restreint l’application de ce texte puisque la preuve qu’une deepfake a été commise par un professionnel doit être apportée. Or cette circonstance est loin d’aller de soi. Ainsi, il peut être attendu et regretté que cette réglementation soit peu appliquée malgré un champ d’application étendu⁴.

⁴ L’AI Act s’intéresse aux personnes, aux objets, aux lieux, et autres entités ou événements existants.

En France, l'obligation de transparence visant les contenus représentant des personnes sera probablement indirectement rendue nulle par l'article 226-8 du Code pénal, la seule mention de la nature algorithmique du contenu permettant d'éviter la poursuite de son auteur, qu'il soit professionnel ou particulier. Cette obligation de transparence restera effective dans les domaines qui ne sont pas couverts par les dispositions nationales.