

Deepfakes & Algorithmes

Menace ou Opportunité ?

Novembre 2020

(club) prAxis

12 propositions

pour repositionner l'Europe comme expert sur les algorithmes de génération de contenu et leader dans la lutte contre les *deepfakes*

Résumé

Les *deepfakes* apparaissent désormais comme un outil de manipulation dont **l'impact sur la société est encore peu maîtrisé**. Leur existence et utilisation soulèvent de nombreuses questions d'ordre **légal** et **éthique**. Mais, au-delà des lois et règles de gouvernance susceptibles d'être mises en place, un problème fondamental demeure, celui de **l'incapacité à détecter un *deepfake***. Alors que la technologie évolue, **il s'avère de plus en plus compliqué d'identifier un faux**. Développer une connaissance européenne des outils de détection de faux et aussi de reconnaissance du vrai apparaît urgent.

Face à l'ensemble de ces enjeux, le dernier rapport Praxis, *Deepfakes & Algorithmes*, formule douze recommandations autour de **quatre grand axes stratégiques** :

- Faire de l'Europe **un leader dans la lutte contre les *deepfakes***
- Renforcer la **responsabilité des plateformes** au niveau européen
- Construire **un environnement réglementaire adapté** à une lutte efficace contre les *deepfakes*
- **Protéger les citoyens** de l'impact des *deepfakes*

Télécharger le rapport complet:

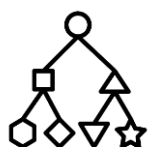
<https://www.clubpraxis.com/publications>

A propos du Club Praxis

Le Club Praxis est un think tank français basé à New York qui rassemble des cadres expatriés bénéficiant de leur expérience à l'étranger pour contribuer au débat en France et l'éclairer de leur expertise internationale. Créé en 2007, sous l'impulsion de plusieurs dirigeants français, dont Henri de Castries, il assume sa position d'outsider dans la vie publique française. Il cherche à promouvoir des idées neuves et innovantes pour le renouveau des institutions et du système économique.

1er Axe

Faire de l'Europe un leader dans la lutte contre les *deepfakes*



1. Soutenir l'amélioration et le développement des techniques de classification de contenu, vrai, faux ou suspect (*fact-checking, watermarks, algorithmes, blockchain*), tout en **définissant des standards européens**, pour accompagner la population dans l'identification de contenu fiable.

2. Investir dans la recherche scientifique sur les algorithmes de génération de contenu, pour être mieux à même de les combattre mais aussi être capable de les utiliser pour des applications positives pour la société.



3. Favoriser les partenariats européens, mutualiser les ressources pour se donner les moyens de rattraper l'avancée de la Chine, de la Russie et des Etats-Unis dans ce domaine.

2ème Axe

Renforcer la responsabilité des plateformes au niveau européen



4. Renforcer la responsabilité encourue par les plateformes (par exemple au sein du Digital Services Act). Imposer à chaque plateforme de mettre en place une structure de gouvernance interne sur la régulation de son contenu. Mettre chaque organe de gouvernance sous supervision européenne en s'inspirant de la régulation mise en place dans le domaine bancaire à la suite de la crise de 2008.

5. Inciter les plateformes à mettre en place des outils pour diminuer l'impact des *deepfakes*

Des sanctions telles que :

- Une pénalité pour les auteurs publiant du contenu faux ou malveillant de manière non explicite.
- Une mention claire mettant en garde l'utilisateur que la publication n'a pu être vérifiée, voire même qu'elle est suspecte.
- Une limitation du nombre de destinataires d'une publication.
- Une limitation des transferts d'une publication par utilisateur.



6. Imposer aux plateformes de partager avec les autorités judiciaires les informations qu'elles détiennent sur les auteurs présumés de *deepfakes*, dans le cadre d'enquêtes pénales, sans que les plateformes ne puissent leur opposer de motif légitime de refus.

3ème Axe

Construire un environnement réglementaire adapté à une lutte efficace contre les *deepfakes*



7. Favoriser l'accès au droit des victimes de *deepfakes*, d'usurpation d'identité ou de harcèlement (aux voies de recours et à l'information de leurs droits) par le développement et la visibilité des **structures d'accompagnement (d'accueil physique ou plateformes étatiques).**

8. Renforcer les sanctions pénales existantes contre les auteurs de contenus de *deepfakes* et les **réévaluer** en prenant en compte le **nouvel environnement technologique et digital** où la création de *deepfakes* est rendue possible.



9. Renforcer au niveau étatique mais surtout supra-étatique (la transmission d'informations ne connaissant pas de frontières) la responsabilité des plateformes.

- **Développer une organisation supranationale de cyber-coopération policière et judiciaire centrée sur la matière afin de faciliter les échanges d'informations entre les Etats.**
- Permettre dans certains cas exceptionnels, **lorsqu'ils diffusent un contenu clairement illégal, d'ordonner la fermeture définitive d'un site**, d'un profil, d'un compte, d'une page, d'un groupe, sans que le site attaqué puisse se défendre
- **Limiter et quantifier légalement le délai de retrait d'un contenu illicite** signalé aux plateformes sous peine de sanctions financières lourdes
- Obliger les plateformes à **vérifier** les données d'identification lors de la création d'un compte et, en **l'absence d'identification réelle de l'auteur, fermer le compte sans délai**
- Imposer aux hébergeurs situés hors de France de désigner un représentant en France qui assume la responsabilité de l'hébergeur qui n'aurait pas retiré dans ce délai légalement fixé un contenu signalé comme illégal
- Assortir systématiquement les sanctions financières infligées aux plateformes d'une mesure de publicité



4ème Axe

Protéger les citoyens de l'impact des *deepfakes*



10. Faire de la lutte contre les *deepfakes* un objectif européen pour 2021. Mettre en place des **campagnes de sensibilisation** pour expliquer de manière pédagogique les *deepfakes*, leurs utilisations et les risques associés en termes de détournement des images, de la voix et des vidéos, tout en insistant sur les applications possibles en manipulation d'opinion.

11. Renforcer la responsabilité encourue par les auteurs. Imposer à chaque auteur de signaler tout contenu *deepfake*.



12. Impliquer les employeurs dans la lutte contre les *deepfakes*. Les inciter à **adopter des chartes éthiques**, à informer leurs employés, clients, fournisseurs et partenaires de leurs bonnes pratiques éthiques sur le sujet. **Encourager la mise en place de formations autour des *deepfakes*.**

A propos des auteurs

Supervision

Emilie Rannou est diplômée de l'ENSAE et titulaire d'un master 2 en mathématiques appliquées de l'Université Paris Dauphine. Senior Manager chez Ekimetrics, elle conseille différentes entreprises sur leur stratégie data science et encadre des équipes de consultants data scientists pour implémenter et optimiser de nouveaux modèles. Emilie a travaillé 12 ans dans l'industrie tech et financière, notamment en tant que Executive Director chez JPMorgan Chase à New York et chez Criteo en tant que Senior Product Manager.

Groupe de Travail

Alexis Benichoux est docteur en traitement du signal de l'Université de Rennes. Il a travaillé autour des méthodes d'apprentissage appliquées au traitement du signal sonore. Alexis est également diplômé du master d'apprentissage de l'ENS Cachan. Actuellement ingénieur de recherche chez Yubo, il développe des algorithmes de reconnaissance d'image et de la recommandation de contenu.

Rémi Forgeas est responsable du France Business Services Center pour un cabinet d'audit et de conseil international. Basé aux Etats-Unis depuis plus de vingt ans, il assiste des groupes français dans leur développement aux Etats-Unis. Il est diplômé de Sciences Po Paris et possède une maîtrise en économétrie.

Simon Gaillard est consultant en protection des données. Basé à New York, il est spécialisé dans le respect de la confidentialité en ligne et aide les entreprises sur leur stratégie de mise en conformité avec le RGPD et le CCPA. Diplômé de Sciences Po Paris, il conseille aussi les organisations internationales sur les politiques publiques, les nouvelles technologies et les questions de protection des données.

Jérémie Mary est docteur en informatique de l'Université Paris-Sud. Il est maître de conférences, habilité à diriger des recherches à l'université de Lille et est actuellement en détachement en tant que chercheur chez Criteo. Il est membre associé de l'équipe Inria SequeL qui est spécialisée dans la prise de décision séquentielle. Il a gagné et organisé 3 challenges pour les systèmes de recommandation dans des conférences majeures de machine learning utilisant des données Yahoo!, Adobe et Twitter (ICML'11, ICML'12, RecSys'14). Son travail de recherche actuel est orienté autour des séquences de recommandations et l'utilisation des modèles génératifs pour les systèmes de recommandation.

A propos des auteurs

Minh Trinh est diplômé de l'Ecole Polytechnique et docteur en économie de l'Université Harvard. Il est consultant basé à New York et est spécialisé dans les méthodes quantitatives et de machine learning appliquées à l'économie et la finance. Il est l'auteur d'un ouvrage sur la gestion des projets en intelligence artificielle. Il est membre de l'American Finance Association et du CFA Institute.

Gabriel Turinici est un ancien élève de l'ENS Ulm et membre de l'Institut Universitaire de France. Il a été chargé de recherche en mathématiques et informatique à l'INRIA jusqu'en 2005 et exerce comme professeur à l'Université Paris Dauphine - PSL. Il est spécialisé dans les algorithmes de simulation, avec des applications en physique quantique, économie, finance et médecine. Il est l'auteur de plus d'une centaine de publications scientifiques et membre du comité de rédaction de plusieurs journaux.

Emilie Waxin est avocate au Barreau de Paris depuis 2010. Après avoir exercé pendant 8 ans en France et au Luxembourg, elle a créé sa propre structure, WE Avocats, en juin 2017. Elle exerce tant en conseil qu'en contentieux en droit des affaires et plus particulièrement en droit social, droit commercial et droit de la propriété intellectuelle. Elle enseigne également le contentieux au sein de l'Ecole de Formation du Barreau de Paris (EFB) et fait partie de plusieurs associations comme l'AFFJ (Association Française des Femmes Juristes).

Club Praxis

Yann Coatanlem est président du club Praxis. Il est diplômé de l'Ecole Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble (ENSIMAG) et du mastère de finance d'HEC, il est directeur de recherche dans une institution financière, président de la Maison Française de New York University, président de l'American Foundation of Paris School of Economics, Conseiller du commerce extérieur et administrateur de la European-American Chamber of Commerce.

En collaboration avec:

Ekimetrics.