Avril 2024

Contribution de KeeeX aux Etats Généraux de l'Information

KeeeX est un éditeur de logiciels français spécialisé dans la confiance numérique et pionnier du contenu vérifiable. La société, qui fêtera ses 10 ans cette année, valorise un procédé universel et breveté issu de la recherche publique française permettant de facilement authentifier et rendre vérifiables tous types de contenus.

Ce nouveau standard de la confiance numérique prend la forme d'un tatouage invisible et permet d'équiper chaque fichier d'un passeport numérique inviolable et facilement vérifiable. Le lecteur d'un contenu peut ainsi en un clic vérifier l'intégrité du fichier, sa provenance, sa date ainsi que d'autres éléments de contexte : copyright, droit de d'utilisation, lieu de prise pour une image, preuve d'existence sur un registre électronique, liens vers d'autres contenus, etc.

Plus de 500 sociétés réparties dans une soixantaine de pays dans le monde utilisent nos solutions. KeeeX est notamment leader sur le marché français du communiqué de presse vérifiable avec des sociétés comme Société Générale et Enedis et des éditeurs de logiciels RP comme Augure et ePressPack qui utilisent notre technologie pour réduire le risque de Fake News et de Deepfakes.

La présente contribution vise à informer les groupes de travail de l'existence d'une solution souveraine permettant de certifier et rendre vérifiables les contenus numériques (particulièrement les groupes de travail 1/ "espace informationnel et innovation technologique" et 4/ "souveraineté et lutte contre les ingérences étrangères") et à inviter les différentes parties prenantes à participer à la mise en place de ce nouveau standard de fait.

La nécessité de certifier de l'information

La quantité de données créées chaque année et leur vitesse de diffusion augmentent de manière exponentielle. Cette surabondance d'informations sature les capacités cognitives des individus qui ne savent plus en quelle information ils peuvent faire confiance. L'essor de l'IA Générative renforce ce malaise et rend accessible au plus grand nombre la génération de contenus de plus en plus réalistes, accroissant ainsi les risques de désinformation. Une analyse[1] montre que l'IA Générative a créé plus de 15 milliards d'images en 2023 alors qu'il a fallu 150 aux photographes, depuis la première photographie prise en 1826, pour atteindre cette barre symbolique. Cette surabondance

[1] https://journal.everypixel.com/ai-image-statistics

de contenus fait que l'œil humain n'est plus capable de discerner le vrai du faux. Le World Economic Forum dans son rapport 2024 sur les risques mondiaux[2] classe d'ailleurs la mésinformation et la désinformation par IA comme le risque n°1 sur les 2 prochaines années, avant le climat.

Lutter contre les menaces que représentent la désinformation, la mésinformation et les deepfakes pour la démocratie et les entités est donc un impératif. Il devient urgent de garantir la confiance et la vérifiabilité des contenus en ligne.

Or la seule vérification a posteriori de contenus pour détecter les produits d'IA Générative n'est pas viable : impossibilité d'apporter une réponse 100% fiable, processus long et souvent énergivore, et évolution des modèles d'IA Générative eux-mêmes.

Signer numériquement les contenus avant leur diffusion s'impose comme la seule solution viable, protégeant les créateurs et offrant la possibilité aux lecteurs de retrouver confiance dans l'information. Le passeport numérique du contenu apparait donc comme la solution idéale pour permettre aux créateurs de se protéger et aux lecteurs de vérifier différents éléments propres à leur contenu.

Le procédé universel KeeeX[3] protège et rend vérifiable pour chaque contenu, quel que soit son format :

- Son intégrité, prouvant que le fichier n'a pas été modifié depuis sa protection
- Sa provenance, indiquant l'identité de son auteur
- Sa date, donnant une indication précise sur la date de création du fichier
- Son existence à une date donnée, permettant d'obtenir une preuve d'antériorité
- Son copyright, donnant la possibilité à son auteur de se protéger et d'indiquer qu'il ne souhaite pas que le contenu serve pour le pré-entrainement d'une IA
- Toutes autres métadonnées utiles elles-mêmes certifiées et inviolables

Pour les Auteurs, Photographes et Journalistes

La certification des contenus par leurs auteurs peut se faire facilement au plus proche de la source via un logiciel ou un service web à utiliser. En plus de garantir l'intégrité du contenu, ces derniers vont pouvoir les signer numériquement avec leur identité numérique (personne physique ou personne morale) et obtenir des preuves de date fiables et précises leur permettant de prouver l'antériorité du fichier.

Chaque fichier possède une empreinte numérique unique que KeeeX transforme en un nom prononçable et indexable, facilitant ainsi les recherches et la désignation du contenu. Ce nom unique pourra également apparaitre dans le passeport numérique des différentes versions suivantes de contenu (photographies avec différentes résolutions, images avec des copyrights différents, etc.) pour des besoins de traçabilité et de recoupements d'information.

---

[2] https://www.weforum.org/publications/global-risks-report-2024/in-full/global-risks-2024-at-a-turning-point/

[3] https://keeex.me

Auteurs, photographes et journalistes ont donc à leur disposition un outil universel et multiformat pour protéger et rendre vérifiable l'ensemble de leurs créations.

Pour les Lecteurs

La vérification d'un contenu est également facile et peut se faire via une interface (module de vérification sur un site web ou une application, plugin de navigateur) ou automatiquement via un micro-service. La véracité du contenu peut quant à elle être vérifiée avant ou après keeexage par un fact-checking professionnel ou communautaire qui engage lui-même sa réputation de façon vérifiable.

La qualité et la réputation du signataire est donc primordiale, le lecteur accordera plus de confiance à un contenu intègre signé par une source qu'il connait que par un contenu signé par un inconnu. Il apparait néanmoins nécessaire de permettre à chacun de protéger les fichiers dont il est l'auteur.

L'équilibre entre liberté d'expression et lutte contre la désinformation est maintenu, chaque lecteur pouvant vérifier la provenance et la date de l'information.

Pour les IA

L'essor de l'IA Générative a porté de graves atteintes à la propriété intellectuelle de beaucoup d'individus et d'entités. En effet, les modèles dits LLM ont besoin d'un très grand nombre de données pour leur pré-entrainement. Certaines sociétés n'hésitent donc pas à aspirer l'ensemble des données présentes sur le web, au détriment de leurs ayants-droits.

Il est donc nécessaire d'offrir la possibilité à toute entité ou individu de prouver la paternité et l'antériorité des leurs créations et d'indiquer qu'ils ne souhaitent pas que leur contenu soit utilisé par une intelligence artificielle.

A l'inverse, les entreprises technologiques vont devoir rendre plus transparentes les IA qu'elles développent en apportant notamment des preuves sur les jeux de données qui ont été utilisés pour entrainer leurs modèles. Cela devient facile avec KeeeX car chaque fichier contient son identifiant sous une forme textuelle simple exploitée par les modèles d'IA. Dans un monde où tous les contenus possèdent leur propre passeport, il devient plus aisé d'apporter la transparence nécessaire et d'apporter une juste rémunération aux ayants-droits qui souhaitent permettre l'utilisation de leurs contenus.

Conclusion

Face à l'urgence des menaces que représentent la surabondance d'informations, l'essor de l'IA Générative et la désinformation, le passeport numérique du contenu apparait comme une solution viable, universelle et souveraine pour protéger à la fois les créateurs et les lecteurs.