



CEPIC

Centre of the Picture Industry

Guide de bonnes pratiques pour une utilisation responsable et éthique de l'intelligence artificielle dans la réutilisation et la production de contenus visuels

I. Préambule :

Le Contenu Visuel fait face à une Révolution.

L'intelligence artificielle a le potentiel de stimuler l'innovation, de soutenir le développement culturel, la croissance et la prospérité, et donc de porter bénéfice à la société dans son ensemble, y compris aux créateurs.

Toutefois, si elle n'est pas gérée de manière responsable, l'intelligence artificielle peut causer des dommages volontaires ou non : étouffer l'innovation et la croissance et porter dangereusement atteinte aux droits individuels, y compris au droit d'auteur.

Ces dommages incluent en particulier :

- La propagation de fausses nouvelles ;
- La violation des droits des tiers, en particulier du droit d'auteur ;
- L'aggravation des préjugés et des problèmes liés à la diversité ;
- L'entrave à l'innovation durable, en donnant un avantage concurrentiel à des entreprises qui ne respectent pas le droit des tiers.

Si l'on veut que les avantages l'emportent sur les dommages, volontaires ou involontaires, il est important que les entreprises qui fournissent des produits générés par l'IA respectent un certain nombre de principes fondamentaux.

Notre principale préoccupation porte sur la collecte et l'obtention des Données visuelles. Ces données doivent être traitées avec le même soin que des ressources naturelles dont l'exploitation doit être durable et qui doivent pouvoir être renouvelées.

Trente années de développement sur internet ont montré que le système d'octroi de licences d'exploitation est le meilleur moyen de créer un environnement numérique sécurisé dans lequel les créateurs ont la possibilité de poursuivre leur œuvre de création visuelle et de prospérer.

Une IA responsable et durable permettra aux créateurs de créer plus, de produire plus d'IA.

Ce guide de bonnes pratiques est rédigé dans le but de minimiser le risque de dommages lors de :

- 1) La collecte ou de l'obtention de données visuelles existantes dans le but d'entraîner les algorithmes (fouille de textes et de données), ce qu'on appelle couramment le DATASCAPING**
- 2) La création ou production de nouvelles données visuelles à l'aide de l'intelligence artificielle générative (GenAI)**

Une obligation de responsabilité s'impose à l'ensemble de la chaîne de valeur, depuis l'utilisation de données existantes aux fins d'entraîner l'algorithme jusqu'à la publication d'un contenu visuel final.

Ce Guide vise à répondre aux questions des investisseurs, des producteurs de contenus visuels générés par l'IA et à leurs utilisateurs.

Ces bonnes pratiques concernent les contenus visuels créatifs et éditoriaux en reconnaissant les spécificités de ces deux secteurs.

II . Définitions :

Dans le présent document :

- **Les Données** sont définies comme des informations dérivées de supports visuels, y compris leurs métadonnées (c'est-à-dire les données sur les données), qu'elles soient intégrées, liées, attachées au fichier principal ou sous toute autre forme.
- **La fouille de données est définie comme le processus de collecte et de sourçage de données**, c'est-à-dire le processus d'obtention de données visuelles (input data) de manière systématique à partir de nombreuses bases de données différentes afin d'entraîner des algorithmes qui seront utilisés soit pour modifier d'autres contenus visuels, soit pour produire de nouvelles données (output data), y compris le contenu d'IA générative. Les données visuelles sont généralement collectées sur des sites web ouverts et librement accessibles.

- **La production de données** est définie comme la création d'un contenu visuel qui peut inclure tout ou partie d'un contenu visuel déjà existant.

Les données publiques sont définies comme des informations auxquelles un utilisateur individuel disposant d'un ordinateur personnel, d'une tablette ou d'un appareil mobile peut accéder sans restriction et où les données sont accessibles à tous. Cela ne signifie pas qu'elles font partie du Domaine public. **Les données visuelles protégées par le droit d'auteur peuvent être libres d'accès. Cela ne signifie pas qu'elles sont aussi libres de reproduction et de diffusion.**

- **Le contenu visuel génératif est** défini comme comprenant tout élément de média ou tout contenu qui a été soit entièrement généré ou modifié par un réseau neuronal entraîné (IA) à un point tel que la nature originale de l'expression en est sensiblement différente.
- **Les droits des tiers incluent tous les droits de propriété intellectuelle (droits d'auteur, marques, brevets, droits voisins), les droits à la vie privée, y compris les droits biométriques et les droits de la personnalité/de l'image.**

III. Les bonnes Pratiques ou Principes :

Tout au long du processus de génération de l'IA, depuis la collecte/le sourçage des données jusqu'à la production de contenu visuel, modifié ou génératif, les mêmes principes fondamentaux énumérés ici doivent s'appliquer, à savoir :

1. La licéité
2. La transparence
3. la minimisation
4. L'exactitude
5. La sécurité

1. La licéité

La collecte de données disponibles sur l'internet, que ce soit à des fins d'apprentissage des algorithmes ou pour produire des contenus visuels, modifiés ou génératifs, doit se faire dans le respect des lois et traités nationaux et internationaux en vigueur au moment de la collecte, notamment en matière de droits d'auteur, de propriété intellectuelle et de données, y compris la législation sur la confidentialité des données/le **RGPD**, ainsi que les conditions contractuelles des services des sites web récupérés.

Aucune donnée ne doit être utilisée sans consentement.

Notamment :

Les images ne doivent être collectées qu'après de sites qui ont obtenu leurs contenus visuels de manière licite, c'est-à-dire que ceux-ci disposent des licences correspondantes. Cela signifie que les images ne peuvent être affichées publiquement, distribuées de quelque manière ou, d'ailleurs, vendues, que parce que le site dispose de l'autorisation requise.

Les organisations diffusant des modèles et des jeux de données d'intelligence artificielle, en open source ou non, pour entraîner des modèles d'apprentissage d'algorithmes capables notamment de générer des images numériques doivent indiquer sur quelle base juridique ils ont obtenu les données : soit par le biais d'une licence d'exploitation, soit sur la base d'une exception au droit d'auteur qui doit être spécifiée.

Les conditions d'utilisation du site web doivent être respectées. Si ces conditions d'utilisation n'autorisent pas la récupération des données visuelles protégées par le droit d'auteur disponibles sur le site, cela doit être considéré comme une condition suffisante pour que ce contenu soit exclu de l'application de l'exception relative à la fouille de textes et de données en vertu de la législation de l'Union européenne (Article 4.3 de la directive sur les droits d'auteur dans le marché numérique unique 790/2019/ EU).

En cas d'utilisation sans autorisation ou de non-respect des conditions d'utilisation le retrait des contenus/données récupérées indument doit être effectué sur simple demande des ayants droits.

Les producteurs de contenu visuel modifié ou généré grâce à un algorithme d'intelligence artificielle sont tenus à des obligations supplémentaires.

Notamment :

Les créateurs de contenu visuel génératif ne doivent jamais produire de contenu :

- *Qui soit trompeur ou/et qui causent des préjudices à des tiers*
- *Qui diffame une personne en affirmant qu'elle a agi ou s'est comportée d'une manière qui n'est pas la sienne, ou en disant des choses qu'elle n'a pas dites, etc.*
- *Falsifier l'information en diffusant des événements qui n'ont pas eu lieu ou tronquer ces événements afin d'influencer l'opinion publique.*

2. La transparence

La transparence est essentielle car elle assure le bon fonctionnement et la responsabilisation de l'ensemble de la chaîne de production, de la collecte des données à la création générative.

Sans obligation de transparence, l'ensemble des principes listés dans ce document, y compris le premier principe de licéité, ne peuvent s'appliquer. Sans transparence, il n'est pas possible de rémunérer les ayants-droits qui le souhaitent et ceux-ci ne peuvent non plus exercer leur droit de opt-out.

Ainsi l'accès, la collecte et le traitement des données doivent être organisés de manière transparente et simple afin de garantir que les tiers soient en mesure d'exercer leurs droits, si nécessaire. Le producteur de contenu visuel manipulé et a fortiori génératif doit respecter les règles de transparence sur la manière dont le contenu a été créé.

Notamment :

La mémoire du contenu visuel collecté (y compris s'il s'agit de données collectées sur des sites accessibles de manière publique) doit être conservée y compris l'information sur le domaine des sites web fouillés et toute référence permettant d'identifier les données visuelles collectées.

Si un compte d'utilisateur est créé dans le but de collecter des données, ce compte de messagerie doit être géré de manière suivie par l'entreprise ou la personne chargée de la collecte des données.

À des fins de transparence, des informations précises sur la source de tous les contenus, telles que la bande passante des fournisseurs de technologie, l'adresse IP achetée et utilisée doivent être conservées.

La transparence est extrêmement importante pour éviter une utilisation abusive du contenu visuel telle la falsification de l'information, la violation de la confidentialité des données et la violation des droits d'auteur.

Les contenus visuels, modifiés ou génératifs, doivent donc être signalés de manière appropriée, opportune, claire et visible et indiquer dans la mesure du possible le nom de la personne physique ou morale qui les ont générés ou manipulés.

Les producteurs de contenu visuel modifié ou généré grâce à un algorithme d'intelligence artificielle sont tenus à des obligations supplémentaires

Notamment :

S'assurer que le consentement des personnes photographiées dans le but d'apprentissage de l'algorithme a été obtenu de manière transparente et éclairée et conformément au GDPR, par exemple par le biais de formulaires légaux (model release), accessibles sur le site du producteur.

Informez les utilisateurs des capacités, des limites et des risques potentiels du contenu visuel manipulé ou génératif, par exemple sur le site du producteur.

Informez les utilisateurs que le contenu est génératif ou comprend des éléments modifiés, grâce au légendage, aux métadonnées ou à des moyens cryptographiques.

Soutenez et/ou participez à toutes les organisations/initiatives visant à authentifier le contenu visuel par des métadonnées ou d'autres moyens.

Les entreprises doivent documenter leur processus de création de contenu afin de montrer que des humains sont impliqués dans le processus, s'ils souhaitent éviter toute affirmation selon laquelle l'IA a créé l'intégralité du contenu.

3. La minimisation

La collecte de contenu visuel de même que la production de contenus visuel modifié ou génératif doivent s'inscrire dans le cadre et l'objectif d'un projet déterminé.

L'accès légal, la collecte et le traitement des données doivent être limités à ce qui est directement pertinent et nécessaire à la réalisation du projet spécifique. Les données ne doivent être conservées qu'aussi longtemps que nécessaire pour atteindre cet objectif. Dans le cas d'un contenu sous licence, les données ne doivent être conservées que pendant la durée prévue par la licence accordée.

Notamment :

Les données ne doivent être collectées qu'à des fins de formation d'algorithmes à la suite d'une analyse des risques visant à garantir que les données collectées sont nécessaires pour le projet et qu'elles ne causeront aucun des dommages identifiés dans ce document.

Toutes les données collectées qui ne sont pas nécessaires à l'apprentissage d'un algorithme doivent être supprimées, en conservant des informations permettant d'identifier la source.

Si les données collectées sont stockées, elles ne doivent l'être que pour la durée strictement nécessaire pour effectuer l'analyse du contenu et être supprimées une fois que l'utilisation n'est plus nécessaire.

La quantité de données collectées au fil du temps auprès d'une seule source, tant en termes de qualité que de quantité, doit rester minimale et "insubstantielle".

La fréquence des visites de sites web ne doit pas augmenter de manière significative le volume de trafic d'un site web.

Les fichiers images non utilisés doivent être supprimés.

4. L'exactitude

Les données collectées doivent être exactes, complètes et, si nécessaire, mises à jour au sens du RGPD.

L'exactitude de toutes les données utilisées dans l'apprentissage des algorithmes est essentielle car, dans le cas contraire, les biais éventuels peuvent être perpétués ou même amplifiés dans le produit final.

Notamment :

Toutes les mesures raisonnables doivent être prises pour garantir que les données à caractère personnel qui sont inexactes, incorrectes ou trompeuses soient effacées ou corrigées.

L'algorithme d'IA utilisé pour produire de nouveaux contenus visuels ne devrait pas être entraîné sur des données éthiquement suspectes et non obtenues légalement.

Le processus de création des données visuelles doit être réalisé de bonne foi dans le but d'éviter de créer des biais.

Ainsi les biais doivent être traités en s'efforçant d'utiliser un ensemble de données contenant des personnes aussi nombreuses et diverses que possible dans des contextes variés, et en étant aussi inclusif que possible.

5. La sécurité

La sécurité est importante pour éviter toute violation des droits d'auteur, violation de la vie privée et également pour éviter toute utilisation abusive afin d'introduire intentionnellement des biais.

Toutes les données personnelles ainsi que celles protégées par des droits d'auteur collectées dans un processus de fouille de textes et données doivent donc être protégées. Cela inclut les contenus sous licence qui doivent être protégés de manière adéquate contre tout accès non autorisé.

Notamment :

Toutes les mesures techniques et organisationnelles doivent être prises pour protéger les données collectées contre d'éventuelles attaques ou piratages.

L'intégrité et la sécurité des sites web utilisés pour la fouille des données ne doivent pas être altérées ou violées par des activités de rétro-ingénierie.

Toutes les mesures visant à protéger les données collectées contre d'éventuelles cyber-attaques ou piratages doivent être mises en place.

Il ne doit pas être possible de procéder à une rétro-ingénierie des images générées par l'IA afin de révéler les images individuelles sous-jacentes.

Des ajustements rapides (downranking, suppression, retrait) doivent être effectués chaque fois qu'une plainte est reçue quand des contenus visuels intentionnellement trompeurs et nocifs sont distribués sur une plateforme.

Des politiques de conformité internes accessibles au public respectant les restrictions d'utilisation strictes telles que décrites dans le présent document doivent être mises en place.

Toutes les mesures appropriées doivent être prises pour mettre en place des politiques internes qui seront conformes aux lois applicables et aux bonnes pratiques afin de sensibiliser les employés aux obligations légales en matière de fouille de données. Il conviendra de procéder à des révisions régulières de ces politiques.

L'acquisition de licences d'exploitation pour l'obtention des données doit être privilégiée.

Conclusion :

Le CEPIC est convaincu que l'intelligence artificielle, dans le secteur visuel offre de grandes opportunités aux créateurs et producteurs de contenu visuel.

Pour que ces innovations soient positives, il est nécessaire et suffisant de respecter les principes de base décrits ci-dessus.

Ceux-ci existent déjà dans la législation actuelle (RGPD et propriété intellectuelle) et peuvent facilement être appliqués à de nouveaux usages.

Le CEPIC sera un fervent défenseur de ces principes de base.

Le Guide des bonnes pratiques éthiques de l'IA pour une réutilisation et une production responsables des contenus visuels sera régulièrement mises à jour afin que les exemples reflètent l'évolution de la législation, des pratiques et de la technologie.

A PROPOS DU GROUPE DE TRAVAIL DU CEPIC SUR L'IA

Le groupe de travail AI du CEPIC a été créé en mai 2022 dans le but d'établir un guide des bonnes pratiques pour une utilisation éthique de l'IA et de suivre ses débuts dans l'industrie de l'image.

Il a été composé dans une approche internationale, diversifiée et inclusive, regroupant des participants issus de différentes parties du débat. Les participants sont issus de professionnels de la création de contenu visuel, de juristes et de spécialistes de l'IA.

Ce groupe de travail est permanent et évolutif.

Plus d'information sur la composition et l'évolution du groupe de travail IA peut être trouvée sur le site du CEPIC www.cepic.org

À PROPOS DU CEPIC

Le CEPIC représente des centaines de bibliothèques d'images et d'agences représentant des centaines de milliers de photographes dont l'activité principale est l'octroi de licences directes de contenu visuel hors ligne et en ligne. En tant que détenteurs de droits, les bibliothèques d'images et les agences accordent des licences pour toutes sortes d'utilisations commerciales, aux journaux, aux magazines, à la publicité, aux radiodiffuseurs, etc. Les membres du CEPIC s'adaptent continuellement à des solutions technologiques innovantes pour la croissance des entreprises numériques et ont développé des plateformes numériques sophistiquées pour commercialiser le contenu numérique en ligne et fournir un accès numérique aux images. Parmi les membres du CEPIC, on trouve des acteurs mondiaux tels que **Getty Images, Shutterstock, Adobe Stock**, des places de marché de contenu visuel telles que **Alamy** et **Dreamstime**, des bibliothèques de beaux-arts telles que **Bridgeman Images** ou **AKG Images**, des archives historiques telles que **Roger-Viollet**, des agences de presse telles que **Belga** (Belgique), **TT** (Suède), **ANP** (Pays-Bas) ou **l'AFP** (France), des bibliothèques d'images plus récemment axées sur l'IA telles que **VAIsual**, des associations de photographes telles que **l'Union des photographes professionnels (UPP)**, ainsi que des représentants des associations professionnelles européennes **AEAPAF, BAPLA, BLF, SAB, SAPHIR** et **SNAPIG**.