



## INTELLIGENCE ARTIFICIELLE

## Les IA génératives à l'épreuve du droit d'auteur et des droits voisins

# Approche comparée en droit européen et américain

### Données / contenus entrants (1ère partie)

À l'heure où des intelligences artificielles (IA) dites « génératives » sont capables de produire des œuvres à partir d'instructions humaines sommaires, grâce à leur entraînement sur des textes, des images ou des sons préexistants, des avocats français et américains se sont interrogés sur les questions que soulèvent ces technologies à l'aune du droit d'auteur et des droits voisins européens et du copyright américain. Dans une première partie que nous publions dans ce numéro, les auteurs analysent les conditions du recours à des données / contenus entrants, pouvant être protégés par un droit de propriété intellectuelle, pour l'apprentissage de ces IA. Le mois prochain, nous publierons la seconde partie de cet article consacrée à la protection éventuelle des œuvres générées par ces IA.

Depuis son lancement le 30 novembre dernier, ChatGPT est présenté comme une avancée technologique majeure, au potentiel de création et de langage semblable à celui d'un humain. Microsoft s'est dite prête à investir 10 milliards de dollars dans la société à l'origine de ChatGPT, OpenAI, en vue de l'intégrer à son moteur de recherche Bing et de réussir à concurrencer Google. Après l'énorme succès de ChatGPT et d'OpenAI, Google a annoncé sa version : Google Bard. Bard est le service expérimental de chat conversationnel de Google basé sur l'IA. Il est censé fonctionner de manière similaire à ChatGPT, la principale différence étant que le service de Google exploitera toutes les informations disponibles sur le web, alors que ChatGPT aurait été entraîné avec les données disponibles jusqu'en 2021 uniquement.

ChatGPT et Bard font partie des intelligences artificielles (IA) dites « génératives » basées sur des technologies de deep learning, une méthode d'apprentissage qui consiste à entraîner une IA à l'aide d'un corpus de données composé de textes, d'images et de sons préexistants. Une fois entraînée, l'IA sera capable de faire des liens entre les données et de produire des modèles, capables de générer des contenus nouveaux sans intervention humaine.

Ainsi, ChatGPT génère automatiquement du texte, tandis que d'autres IA comme Dall-E ou Midjourney parviennent à créer des images à partir de descriptions textuelles. Il convient donc de distinguer le corpus de données / contenus utilisés pour entraîner l'IA - qui constitue les « données / contenus entrants », des résultats générés par l'IA - qui constituent les « données / contenus sortants ».

Ces potentielles « œuvres générées par IA » sont créées principalement par un logiciel. Elles se distinguent des « œuvres assistées par IA », pour lesquelles l'IA n'est qu'un simple outil au service de l'humain, qui sont susceptibles de faire « apparaître même de façon minime l'originalité qu'a voulu apporter son concepteur »<sup>1</sup> et que la doctrine internationale classe parmi les œuvres protégeables par le droit d'auteur<sup>2</sup>. Les œuvres générées par IA sont plus difficiles à appréhender par le droit d'auteur et les droits voisins, tant du point de vue des « données entrantes » que des « données sortantes ». Elles sont par ailleurs empreintes d'internationalité puisque principalement alimentées à partir de données accessibles sur internet.

L'Union européenne connaît un régime de protection automatique par le droit d'auteur dès la création