Comment protéger les bases de données au cœur d'un projet d'IA ? Analyse à travers 3 décisions juridiques clés

En tant que Conseil en propriété industrielle, spécialisé en électronique et technologies de l'information, ma carrière a été consacrée à la préservation des innovations dans ces domaines en constante évolution. Ces dernières années, ma compétence a été sollicitée de manière récurrente pour l'obtention de brevets liés à l'intelligence artificielle (IA). Cette technologie, dont les contours sont aujourd'hui encore difficiles à définir, représente un défi unique pour le système de brevet traditionnel.

Dans cet article, je vous propose de comprendre la protection les bases de données par le droit *sui generis*, à travers plusieurs décisions qui ont récemment clarifié les moyens de justification d'un investissement substantiel dans la collecte, la vérification et la présentation de données, et ce dans le but de se prévaloir de ce droit *sui generis*.

Base de données, IA... : Bien comprendre le contexte

De nos jours, la plupart des innovations en matière d'IA reposent sur des modèles bien établis et connus de l'industrie comme *Inception*, *ResNet*, *VGG*, *DenseNet*... Par exemple, le désormais réputé modèle VGG 19 pourrait être utilisé pour déterminer comment les lecteurs comprennent cet article sur la base d'une analyse par webcam des expressions faciales. Dans ce cas théorique, le modèle pourrait alors être adapté ou légèrement modifié pour atteindre de très bons résultats pour cette implémentation spécifiquement. Ces modifications peuvent, en théorie, être protégée par le mécanisme du brevet. Cependant, même avec l'obtention d'un brevet, les concurrents pourraient utiliser un modèle alternatif, obtenant des résultats similaires sans pour autant être en position de contrefacteur du brevet.

De plus, lorsque l'on dépose un brevet, la loi impose la fourniture d'une description complète de l'invention concernant l'état de la technique (ou état de l'art). Or, cette description, en tant que partie intégrante du brevet, fait l'objet d'une publication pendant la procédure d'examen et de délivrance. Ainsi, si le brevet est trop spécifique, les concurrents pourront tirer des leçons de cette divulgation et améliorer leurs propres produits moyennant des efforts et des coûts amoindris sans être contrefacteur.

Reconnaissant les limites de la protection par brevet pour des implémentations spécifiques de l'IA, j'ai recherché des mécanismes alternatifs qui pourraient potentiellement offrir une protection à ces projets d'IA. Ainsi, j'ai été amené à réaliser une exploration plus approfondie de la protection des bases de données.

En effet, nous constatons souvent que l'effort et les ressources déployés pour l'acquisition de données dédiées à l'entraînement de l'IA surpassent souvent ceux investis dans l'écriture du code sous-jacent à l'IA. Cette dernière tire sa réelle puissance et sa différence de ces bases imposantes d'apprentissage, formées par la collecte et l'accumulation de données labellisées. La véritable pierre angulaire de l'IA réside donc dans l'énorme besoin de données pour son entraînement, base de données qui nécessitent un investissement important.

Aujourd'hui, les bases de données sont protégées par différents droits de propriété intellectuelle et moyens de protection, notamment :
- le Droit d'auteur (ou Copyright), qui sauvegarde de la structure ou des données stockées dans la base de données ;
- le RGPD, qui protège un refuge à des catégories de données spécifiques ;
- les Contrats, limitant l'accès ou l'utilisation des bases de données ; et
- le droit *sui generis* des bases de données, qui protège les investissements.

Lorsqu'on considère les projets d'IA, le droit sui generis apparaît comme particulièrement intéressant.

Description du droit *sui generis*

En droit européen, le droit sui generis s'applique aux bases de données qui ont nécessité un investissement **substantiel** pour **leur collecte**, **leur vérification** ou **leur présentation**.

Plus précisément, selon l'article 7 de la directive européenne sur les bases de données (1996), le droit sui generis vise à empêcher une **extraction qualitativement ou quantitativement substantielle** d'une base de données lorsqu'un **investissement substantiel a été réalisé**. À titre d'exemple, si une base de données protégée par le droit sui generis est disponible en ligne, il n'est pas possible d'en extraire toutes les données pour créer une nouvelle base de données ou améliorer une base de données existante.

Les règles qui entourent ce droit *sui generis* sont encore en discussion pour être complétées et précisées. Il y a eu cependant quelques affaires importantes qui nous aident à comprendre son fonctionnement.

L'une des jurisprudences est celle de *British Horse Racing Board* contre *William Hill*, devant la Cour de justice de l'Union Européenne.

Dans cette décision, la société B*ritish Horse Racing Board* (BHRB) disposait d'une base de données contenant de nombreux détails sur les courses de chevaux au Royaume-Uni, comme les noms des chevaux, leurs côtes, les résultats et les lieux des courses... Ces informations étaient accessibles au public sur leur site Web. *William Hill*, une société de paris sportifs bien connue au

Royaume-Uni, a utilisé certaines de ces données pour sa propre plateforme de paris en ligne.

BHRB a accusé *William Hill* d'atteinte à son droit *sui generis* par l'extraction et l'utilisation de ses données.

La Cour a déclaré que *BHRB* avait investi beaucoup de temps et de ressources dans la création de sa base de données, mais *BHRB* n'était pas en mesure de prouver que ces investissements visaient directement **la collecte**, **la vérification** ou **la présentation** de la base de données. Par ailleurs, elle considère que le *BHRB* ne peut prouver que *William Hill* ait extrait une partie **qualitativement ou quantitativement** significative de la base de données. De la sorte, *BHRB* n'a donc pas pu obtenir gain de cause sur le fondement de l'atteinte à son droit sui generis.

En France, un différend similaire est intervenu entre Ryanair et Opodo : Ryanair a investi pour construire une grande base de données sur les vols, mais n'a pas réussi à prouver cet investissement lorsqu'*Opodo* a été poursuivi en justice en raison de la récupération du contenu de cette base de données. Dans cette affaire, *Opodo* a fait valoir que *Ryanair* n'avait pas correctement détaillé la **structure de sa base de données**. Le tribunal considère qu'il est impossible de définir l'investissement dans les tâches liées au droit *sui generis* **si l'on ne peut pas expliquer le volume global de travail impliqué dans la construction de la base de données**. Ainsi, le niveau de détail demandé par le tribunal se rapproche de celui de la définition donnée dans un brevet, lequel doit expliquer toutes les étapes impliquées dans la construction d'une IA.

En conclusion, au vu des résultats de ces décisions, il semble difficile de prouver :
- l'investissement, y compris la définition des **tâches globales impliquées dans la construction de la base de données** ; et
- que l'extraction doit être sanctionnée car **qualitativement ou quantitativement significative**.

En raison des standards de preuve très élevés exigés par les tribunaux, le droit sui generis est actuellement peu utilisé comme outil de protection des bases de données. Cependant, des évolutions récentes laissent penser que cela est susceptible de changer.

Nouvelles décisions françaises et leurs conséquences

Le 5 octobre 2022, la Cour de cassation a rendu une décision très intéressante dans la désormais célèbre affaire impliquant *leboncoin.fr* et *Entreparticuliers.com.* Les deux sociétés ont un modèle économique basé sur une place de marché en ligne et la publicité, mais *Entreparticuliers.com a* été accusé d'avoir récupéré toutes les données de la base de données immobilière du site *leboncoin.fr.*
Dans ce cas, *LeBonCoin* a pu prouver tous les investissements réalisés pour **la collecte** et **la vérification** de ses annonces, avec une explication précise du traitement des données depuis le comportement de l'utilisateur sur le site jusqu'à la publication des annonces. En effet, il a été démontré que de nombreuses analyses et traitements sont réalisés pour éviter le harcèlement sexuel, les spams, les erreurs...

En outre, la cour a déclaré que si une entreprise rapporte la preuve d'investissements substantiels dans une partie restreinte de la base de données principale, comme une base de données immobilière, **cette partie peut bénéficier de sa propre protection juridique distincte sur le fondement du droit *sui generis***. Dans cette affaire, leboncoin.fr avait suffisamment investi dans sa rubrique immobilière pour mériter sa propre protection, ce qu'*Entreparticuliers.com* avait extrait en totalité.

Cette décision française explique comment on peut justifier de l'investissement et obtenir une condamnation sur le fondement du droit *sui generis*.

Une autre décision française récente impliquant la société *LaCentrale.fr* a confirmé que l'on peut prouver l'investissement et l'extraction en utilisant la méthode développée dans la jurisprudence *LeBonCoin*.

Ces récentes affaires, nous permettent de mieux comprendre comment justifier les investissements protégeables par le droit sui generis, et d'envisager ainsi de nouvelles stratégies visant à la protection des actifs immatériels de l'entreprise.

En conclusion, bien que cela nécessite effectivement une analyse approfondie de la base de données et des connaissances générales en informatique, il apparaît que le droit *sui generis* est à considérer comme un moyen efficace de protéger les projets d'IA, au vu des investissements substantiels impliqués.

Nous conseillons vivement les entreprises d'anticiper une utilisation potentielle de ce droit sui generis et de réaliser, lorsque les traitements sur les données sont structurés, un audit de leurs bases de données visant à évaluer les investissements réalisés pour la base de données avec l'application du droit sui generis. Cet audit permettrait d'apporter des éléments de preuves solides si une extraction devait être réalisée.

Gaëtan LASSERE, associé Gérant et Mandataire en Brevets Européen chez Laurent & Charras