



NÉGOCIER À L'HEURE DE L'IA

IMPACT DE L'IA SUR LES MARCHÉS DE DOCUMENTATIONS ÉLECTRONIQUE



LES IMPACTS SUR LE PÉRIMÈTRE DU MARCHÉ

□ Contexte: 2eme semestre 2023, 1 an après la sortie de Chat GPT

- Des clauses limitant la portée des usages des outils IA sur les contenus, dans le contexte de bases bibliographiques (Web of Science / Scopus) : exemple **d'interdiction** :

*utiliser les Produits Souscrits en combinaison **avec un outil d'intelligence artificielle** (y compris pour entraîner un algorithme, tester, traiter, analyser, générer des résultats et/ou développer toute forme d'outil d'intelligence artificielle*

- Constat : définition trop générale, interdisant même la fouille de texte et de données (TDM), pour des contenus utilisables précisément à cette fin

LE TDM DANS LE CONTEXTE ESR FRANÇAIS

- ❑ Codifié dans le code de la propriété intellectuelle article L 122-5-3 (Ordonnance n° 2021-1518 du 24 novembre 2021 complétant la transposition de la directive 2019/790)
 - Limite les usages : exclusivement la recherche
 - N'institue pas un droit de mise à disposition des contenus (sauf pour la simple vérification des données) : n'est pas une ordonnance science ouverte
 - N'ouvre pas de perspectives quant au partage des résultats
- ❑ Limitation des usages TDM par un contrat : impossible → sensibilisation, par exemple par l'ajout systématique **des textes législatifs** :
 - *A l'exception de ce qui est prévu à l'article L 122-5-3 du CPI...*

L'ORDONNANCE TDM COUVRE-T-ELLE TOUS LES USAGES IA ?

- ❑ En France, une longue histoire du **TDM** dans la collaboration Abes-Couperin-CNRS : **projet ISTEEX**
 - Contractuellement, obtenir davantage que ce que la loi autorise
 - Structurer des textes pour permettre l'extraction de corpus
 - Usages : analyse, interprétation, calculs de tendances, des relations entre sujets, revue de littérature ++,...
- ❑ Mais les nouveaux outils changent la donne : ChatGPT, Llama, Mistral... → créer du nouveau contenu textuel à part entière, ouvrir de nouveaux horizons :

IA ET PROCESSUS ÉDITORIAL

- ❑ Pour les éditeurs, l'IA générative soulève des **opportunités**
 - Dans le processus éditorial
 - Dans la génération des textes liminaires des publications
 - Dans la mise en place de nouveaux services à partir des contenus
- ❑ Et, pour cette même raison, pose des problèmes en termes d'usages :
 - **Partage** des contenus par les utilisateurs : avec qui, comment ?
 - Génération d'**oeuvres dérivées** : par qui, dans quels contextes ?

CLAUSES IA ET APPROCHE PRAGMATIQUE

- ❑ Un exemple de compromis : **délimiter l'interdiction** :

*utiliser les Produits Souscrits **en combinaison avec un outil d'intelligence artificielle générative** (y compris pour entraîner un algorithme, tester, traiter, analyser, générer des résultats et/ou développer toute forme d'outil d'intelligence artificielle) sauf si cet outil d'intelligence artificielle est utilisé dans un environnement local auto-hébergé **ou sur un système de type cloud uniquement pour l'usage des Utilisateurs autorisés** et qu'il ne partage tout ou partie des Produits Souscrits avec une tierce partie;*

- ❑ Rendre possibles les usages constatés aujourd'hui
- ❑ Ne pas limiter à des finalités de recherche (d'autres besoins possibles)
- ❑ Miser sur le développement de la SO

CLAUSES IA ET CONNAISSANCE DES USAGES

- ❑ **En attendant une réglementation européenne en faveur de l'IA**
 - Cibler les usages indispensables et les rendre possibles contractuellement
 - Prudence : ne pas considérer que le cadrage du TDM cadre toute l'IA
 - Ne pas se tromper d'objectif : négocier la SO / négocier les usages d'outils
- ❑ **Expérience TDM : les besoins ESR risquent d'être plus importants que la définition du cadrage légal → négocier des acquis supplémentaires (exemple : pilotage recherche)**

EXEMPLE DE CLAUSE INTÉGRÉE DANS LES CONTRATS

A l'exception de ce qui est expressément prévu dans le présent Contrat ou autorisé par écrit par l'éditeur et hormis les usages de TDM autorisés par l'article L122-5 du Code de la Propriété Intellectuelle complétant la transposition de la directive 2019/790 du Parlement européen, l'Abonné et ses Utilisateurs Autorisés ne peuvent pas:

créer une quelconque œuvre dérivée et/ou service (y compris ceux résultant de l'utilisation d'outils d'intelligence artificielle), à partir des Produits Souscrits, à l'exception de ce qui est nécessaire pour les rendre perceptibles sur les écrans d'ordinateurs pour les Utilisateurs Autorisés;

utiliser les Produits Souscrits en combinaison avec un outil d'intelligence artificielle générative (y compris pour entraîner un algorithme, tester, traiter, analyser, générer des résultats et/ou développer toute forme d'outil d'intelligence artificielle) sauf si cet outil d'intelligence artificielle est utilisé dans un environnement local auto-hébergé ou sur un système de type cloud uniquement pour l'usage des Utilisateurs autorisés et qu'ils ne partagent tout ou partie des Produits Souscrits avec une tierce partie;



Négocier les clauses IA : “passé, présent, futur”

Présentation des travaux menés par Couperin pour intégrer l'IA dans la trajectre des négociations de documentation électronique

Approche juridique de l' IA : un continuum d'outils, du big data au génératif ?

« L'intelligence artificielle est un procédé logique et automatisé reposant généralement sur un algorithme et en mesure de réaliser des tâches bien définies. Pour le Parlement européen, constitue une intelligence artificielle tout outil utilisé par une machine afin de « *reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité* ». Plus précisément, la Commission européenne considère que l'IA regroupe :

- les approches d'apprentissage automatique ;
- les approches fondées sur la logique et les connaissances ; et
- les approches statistiques, l'estimation bayésienne, et les méthodes de recherche et d'optimisation. »

Source CNIL

A person in a blue suit is holding a European Union flag. The background is a light blue gradient with a faint pattern of yellow stars.

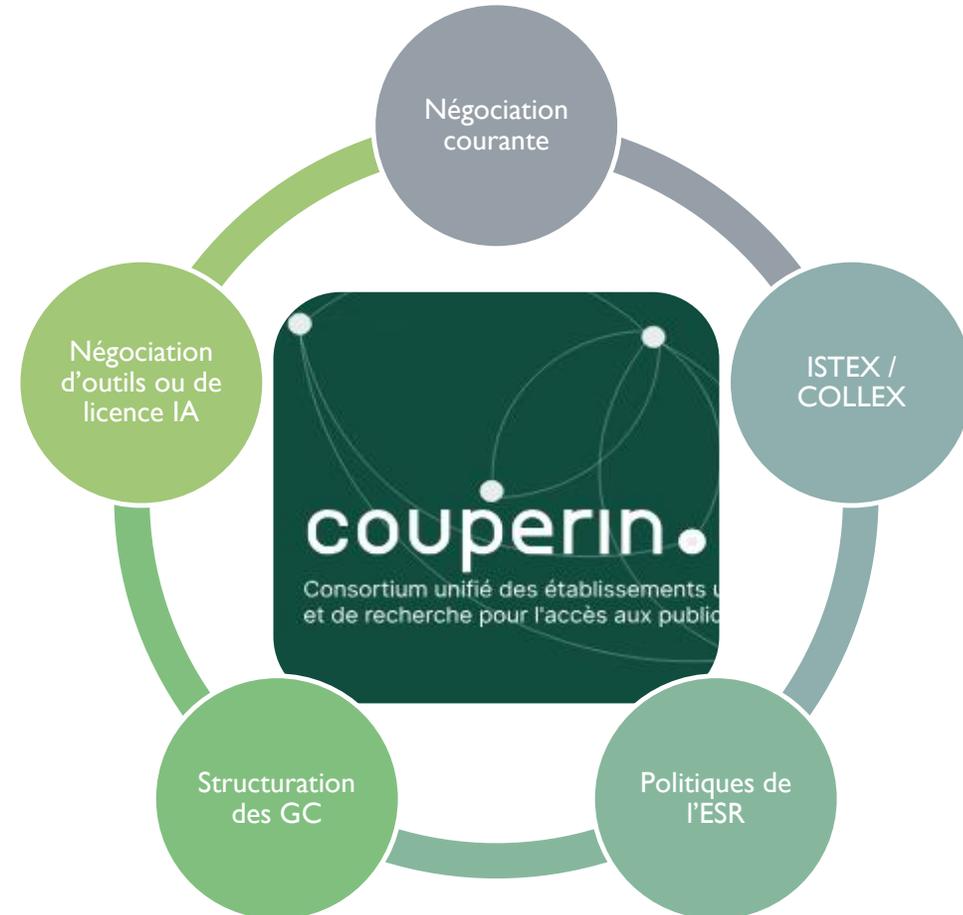
'Artificial intelligence system' (AI system) means a system that is designed to operate with a certain level of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of human-defined objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions , influencing the environments with which the AI system interacts.

Source Parlement UE – AI Act

L'IA : quels enjeux pour le consortium ?

L'IA impacte directement l'ensemble du champ des négociations :

- Négociation courante : adaptation aux demandes du fournisseur sur la base d'éléments partagés ; obligation de rester dans un cadre juridique « applicable »
- Négociation nationale : lien avec ABES et relative harmonisation des clauses
- Négociation ISTEK/ Collex : mise à disposition de corpus, perspective de fournir une gamme de services identiques pour les ressources
- Négociation dans le cadre de son rôle « d'opérateur » IST : défendre des éléments de régulation globale
- **A VENIR ?** Négociation d'outils IA : à construire avec les membres (évaluation des outils, modèle économique, articulation avec ESR)



L'IA : quels enjeux pour les membres de Couperin et la fourniture de documentation électronique?

Principes de structuration :

- Besoin d'éléments partagés et de lignes rouges au sein de Couperin
- Ces besoins sont définis par des usages (recherche, enseignement, documentaire)
- Éléments consensuels
- Compréhension des besoins des fournisseurs selon leur nature



Le GT IA de Couperin : une structure ad hoc pour comprendre les outils et proposer des éléments de cadrage des négociations

- Groupe réuni de janvier à avril 2024, comprenant l'ensemble des métiers concernés : chercheurs spécialisés, DPO, juristes, bibliothécaires
- Alimenté par les travaux de l'ABES
- Représentatif de l'ESR : organismes (CEA, INRIA, INRAE), universités (UPSaclay, AMU, Bordeaux-Montaigne), écoles (Mines Paris-PSL), ABES
- CONTEXTE : multiplication d'apparition de clauses d'interdiction de l'usage de l'IA dans les licences
- OBJECTIFS : Fournir des éléments de compréhension technique et juridique, définir un cadre partagé d'éléments qui puisse être fourni aux négociateurs, fournir une première version de clause type

Proposition de clause type en juin 2024

L'usage de intelligence artificielle sur les données détenues par l'AFNOR ou pour lesquelles elle détient les droits de distribution n'est autorisé qu'à des fins exclusives de recherche .

Par Intelligence artificielle, on entend les techniques et outils visant à imiter les facultés cognitives humaines telles que l'apprentissage, le raisonnement, la résolution de problèmes, la compréhension du langage naturel, la reconnaissance de motifs et la prise de décisions.

L'utilisation de l'intelligence artificielle à fins de recherche doit s'effectuer localement ou dans une application hébergée sécurisée.

Il est interdit dans tous les cas de :

- reproduire et divulguer les contenus générés*
- d'avoir un usage pouvant entraver la bonne exploitation par l'AFNOR*
- donner accès à une société commerciale au contenu*

Pour les négociations courantes : des briques à intégrer selon les demandes des fournisseurs

- **DEFINIR SYSTEMATIQUEMENT LES TERMES**

Génératif : « Traitement permettant de produire par interaction des contenus sous un arrangement original, à l'exception des systèmes de transformation de contenu (résumé, traduction automatique, amélioration de qualité d'images, amélioration grammaticale et stylistique...) »

Entraînement : « Processus de production d'un système d'IA par analyse statistique de contenus/données sources »

Agent conversationnel : « Système automatique permettant d'interagir avec une machine au moyen du langage naturel »

- **DISTINGUER USAGE RECHERCHE / ENSEIGNEMENT EN DEMANDANT UNE EXCEPTION**
- **PROPOSER UNE BRIQUE PARTENARIALE**
- **INTEGRER DES CAS D'USAGE**
- **NEGOCIER A MINIMA LE DROIT D'UTILISER L'IA DANS UN ENVIRONNEMENT FERME ET SECURISE**
- **DEUX LIGNES ROUGES** : RIEN NE DOIT METTRE EN PERIL L'USAGE TDM / IMPOSSIBILITE DE FOURNIR A DATE DES PLANS D'ASSURANCE SECURITE

Enjeux de régulation constatés au cours des échanges du GT

Enjeux	Argumentaire
Transparence (Mise en cohérence avec le Règlement européen)	<ul style="list-style-type: none">• Les IA sont une évolution naturelle des moteurs de recherche• Enjeu de reproductibilité• Transparence concernant les flux financiers générés par l'IA• Divulgence des modèles entraînés sur le corpus éditeurs
Equité vis-à-vis des auteurs	<ul style="list-style-type: none">• L'utilisation de l'IA doit être intégrée dans le contrat d'auteur. Rémunération de l'auteur
Responsabilité	<ul style="list-style-type: none">• L'IA n'est pas un sujet en soi, mais se réduit in fine à la question de la responsabilité académique• Recevabilité de recherches menées dans un cadre strictement prédictif• Garanties concernant la bonne utilisation des données fournies par les éditeurs (revente)

Des premiers éléments globaux de réflexion : préserver un cadre « Qualité recherche »

- En cas de rémunération des éditeurs grâce à de l'exploitation ou revente de solutions IA, quelle rémunération pour les auteurs ?
- Ouverture des LLMs et RAGs mis en œuvre
- Mise à disposition des corpus enrichis automatiquement par IA générative
- Conclusions qui rejoignent les préconisations du Règlement européen sur l'Intelligence artificiel du 13 juin 2024 : respect droit d'auteur et droit voisin, publicité de « résumé suffisamment détaillé »
- Pour la réflexion concernant le début de mise en œuvre, voir le récent rapport du CSPLA publié le 11 décembre 2024 : « **Rapport d'Alexandra Bensamoun sur la mise en œuvre du règlement européen établissant des règles harmonisées sur l'intelligence artificielle (RIA)** »
- **Pour l'ESR, les enjeux liés à la préservation de la qualité de la recherche rejoignent cet objectif de transparence, et gagneraient sans doute à être intégrés à la Science ouverte**

L'IA pour Couperin : et maintenant ?

1. En 2024/2025, important travail sur les clauses au fur et à mesure de leur soumission dans le cadre des négociations et/ou dans le cadre de l'ICOLC

Très grande hétérogénéité des rédactions initiales et de leur qualité :

- Intégration du TDM et identification à l'IA
 - Clause impliquant la responsabilité juridique de Couperin
 - Le plus souvent, clause d'interdiction sauf pour certains usages ou cadres d'usage (reconnaissance des environnements sécurisés, end user)
 - Blocage sur la divulgation des corpus produits, parfois ouverture concernant la mise à disposition des LLMs
 - Distinction émergente entre IA tierce et IA propriétaire
2. Par ailleurs, Couperin souhaite initier la réflexion sur les aspects autres que juridiques, et en particulier concernant la qualité des moteurs de recherche et des outils génératifs : lancement d'un groupe dédié