

École centrale Casablanca

Projet Intraprendre une spin-off : JURISINNOV'

Cahier de charge

Réalisé par:

LAHRIQUI Adam
EL MRABET Aimane
EL HOUARI Nada
DAHASSI Chaymae
OUANZOUGUI Abdelhak

Encadré par:

Mr. Jean-Pierre Llored (FECC)

Table des matières

Table des matières	1
Table des figures	2
1 Introduction et Contexte	3
2 Problématique	3
3 Solution	4
4 Base de données	6
5 Algorithmes	6
6 Résultats préliminaires	7
7 Business Model	9
8 Perspectives	9
9 Conclusion	10

Table des figures

1	Schéma décrivant la solution adoptée	4
2	Interface de Connexion de notre MVP	5
3	Interface de Recherche de notre MVP	5
4	Le code d'extraction de métadonnées sur gethub Arabert	7
5	Résultat de conversion de fichier par OCR docTR	8
6	Exemple de résumé	8

1 Introduction et Contexte

Dans le domaine juridique, l'accès aux documents de jurisprudence est un aspect crucial pour les professionnels du droit, y compris les avocats, les juristes et les étudiants en droit. Ces documents, généralement disponibles sur les sites web du ministère de la justice, sont souvent longs et présentés dans un format brut et non structuré. Naviguer à travers ces documents étendus pour extraire des informations essentielles, telles que le sujet du document, la décision du juge et les arguments présentés par chaque partie du cas, représente un défi significatif pour les praticiens du droit. L'absence d'un processus rationalisé pour extraire efficacement et résumer l'information des documents de jurisprudence crée un vide dans le paysage technologique juridique.

Conscients de ce défi, notre projet vise à tirer parti des avancées en matière de traitement du langage naturel (NLP) et d'intelligence artificielle (IA) pour transformer la manière dont les professionnels du droit interagissent avec les documents de jurisprudence. En automatisant l'extraction des données essentiels, tels que le sujet du document, la date et le tribunal émetteur, et en générant des résumés concis, nous avons l'intention d'améliorer l'accessibilité et l'efficacité dans la recherche juridique et la prise de décision.

L'opportunité sur le marché pour notre solution est substantielle, étant donné la dépendance généralisée aux documents de jurisprudence dans la pratique juridique. Les cabinets d'avocats, les avocats individuels, les chercheurs juridiques et les institutions éducatives peuvent bénéficier considérablement d'un outil qui simplifie le processus de digestion de documents juridiques complexes. De plus, le manque de solutions existantes pour la génération automatisée des résumés des documents de jurisprudence positionne notre projet comme un projet innovant dans l'espace technologique juridique.

2 Problématique

Le secteur juridique est confronté à des défis importants en matière de gestion et de récupération de la jurisprudence, caractérisés par des méthodes de recherche laborieuses, des obstacles à l'accès à des informations juridiques actualisées et la complexité de la gestion de vastes ensembles de données juridiques. Ces problèmes entraînent la prolongation des procédures judiciaires, l'augmentation des coûts de recherche et l'accès limité à des précédents juridiques cruciaux. Comment peut-on réduire efficacement ces retards persistants ? Quelles stratégies pourraient être employées pour rationaliser les coûts de recherche et améliorer l'accès à la jurisprudence ?

3 Solution

Notre solution englobe un flux de travail complet qui commence par une phase de prétraitement. Au cours de cette étape initiale, la technologie de reconnaissance optique de caractères (OCR) est utilisée pour convertir les documents de jurisprudence scannés en fichiers texte. Ensuite, l'architecture asynchrone, composée de deux codes, prend le relais. Le premier code utilise le traitement du langage naturel (NLP) pour extraire des informations cruciales telles que la date du document, le sujet et le tribunal émetteur, puis génère un résumé automatique qui inclut la décision du juge et les arguments de chaque partie. Ces métadonnées résultantes sont efficacement stockées dans une base de données dédiée. Le deuxième code facilite l'accès de l'utilisateur à la base de données via une barre de recherche et un algorithme de similarité. En intégrant les technologies OCR et NLP, notre solution rationalise les processus de conversion et d'extraction, offrant aux professionnels du droit un moyen efficace de naviguer et de comprendre les documents de jurisprudence.

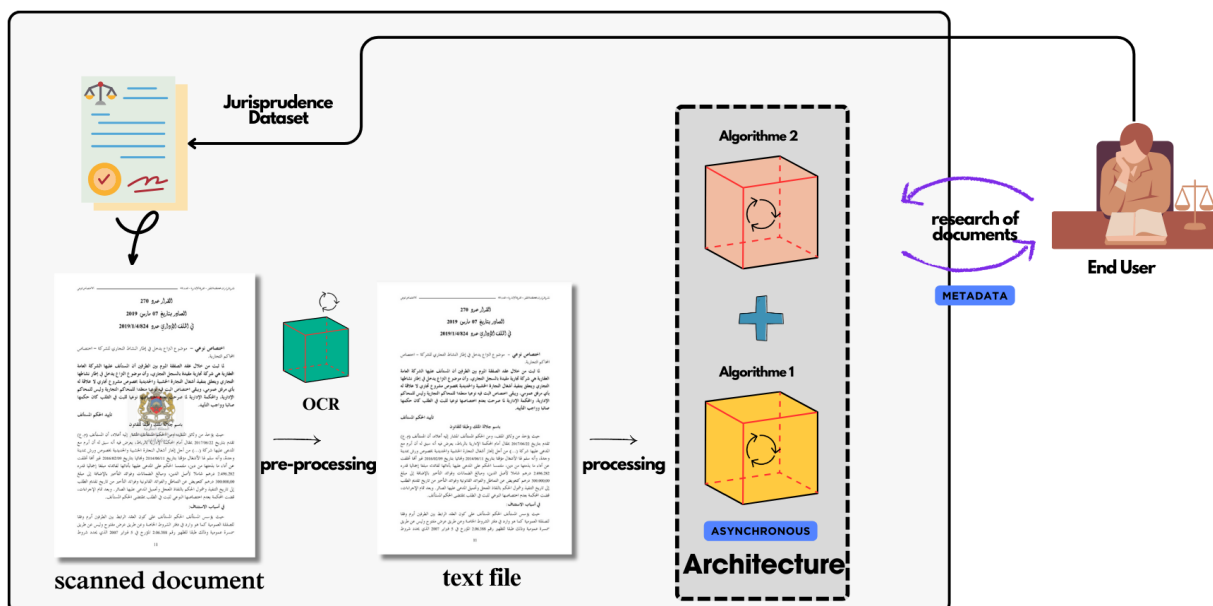


FIGURE 1 – Schéma décrivant la solution adoptée

Notre Minimum Viable Product (MVP) est une plateforme permettant à chaque utilisateur d'entrer ses informations de connexion et d'accéder ensuite aux diverses données et services proposés par notre plateforme.

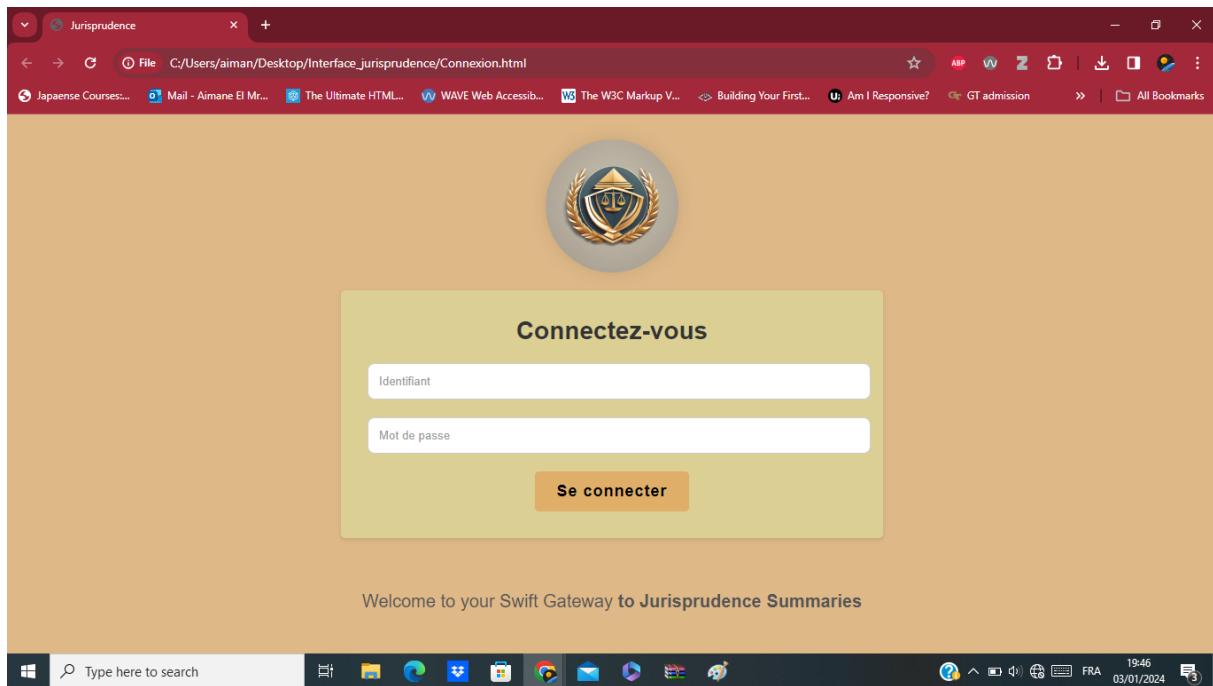


FIGURE 2 – Interface de Connexion de notre MVP

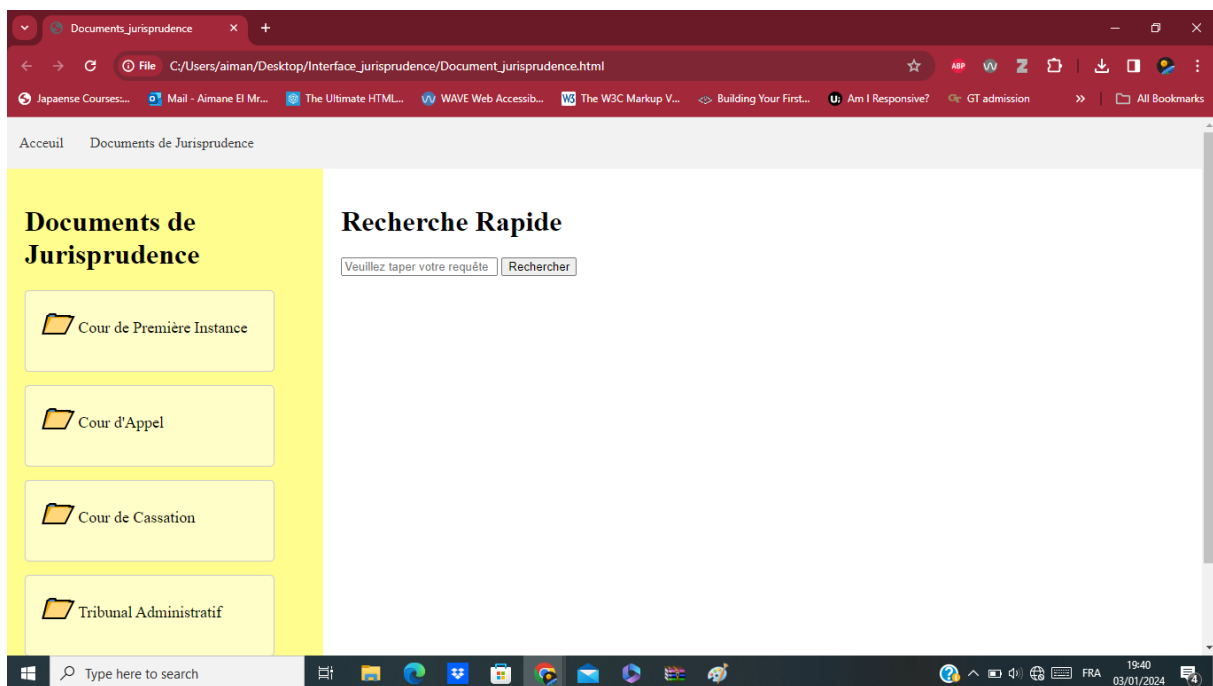


FIGURE 3 – Interface de Recherche de notre MVP

4 Base de données

Notre projet s'appuie sur la numérisation de la jurisprudence ainsi que la génération automatique de résumés. L'objectif principal est de rendre la jurisprudence plus accessible et facilement consultable pour tous. Pour réaliser ce projet ambitieux, il est essentiel de disposer d'une base de données exhaustive de décisions judiciaires. Dans cette optique, nous avons pris l'initiative de collaborer avec des experts dans le domaine. Nous avons ainsi approché Maître Laraqui qui a généreusement accepté de nous aider. Grâce à son soutien, nous avons pu obtenir un trésor d'informations : les archives judiciaires des vingt-trois dernières années. Néanmoins, la data que nous avons reçue se présente sous forme de fichiers PDF munis d'un filigrane central. Ainsi, une étape préliminaire essentielle pour notre projet est de soumettre ces documents à un processus de reconnaissance optique de caractères (OCR). Ce traitement permettra de convertir le contenu des PDF en texte exploitable pour nos analyses et notre plateforme.

5 Algorithmes

Après avoir reçu, de la part de Madame Laraqui, la base de données sur laquelle nous allons travailler et tester notre solution, la première étape que nous avons dû franchir était le prétraitement de la base de données. Cela était nécessaire en raison de la nature de la base de données, qui consiste en un ensemble de fichiers de jurisprudence scannés, chacun contenant un filigrane au milieu du fichier. En collaboration avec nos camarades de troisième année, qui nous ont accompagnés tout au long du développement de notre solution, il était essentiel de trouver un outil permettant de convertir ces fichiers scannés en fichiers texte que nous pouvons traiter.

La recherche effectuée par nos camarades a révélé que l'outil de reconnaissance optique des caractères (OCR), et plus précisément docTR, était le plus performant en termes de rapidité et d'efficacité. Ils ont constaté une efficacité de 93% et un temps de réponse de 173 secondes. Ensuite, après avoir converti tous les fichiers, vient l'étape principale du traitement qui concerne la partie de l'architecture asynchrone, composée de deux algorithmes fonctionnant simultanément et de manière indépendante.

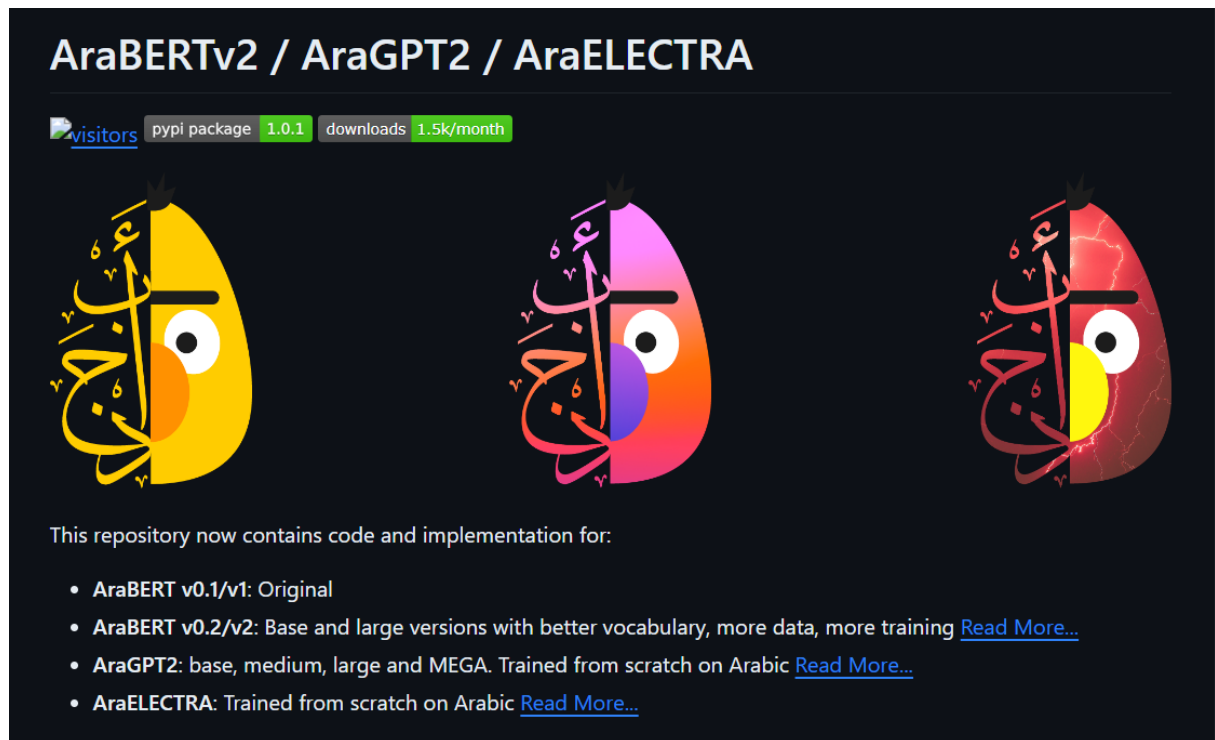


FIGURE 4 – Le code d'extraction de métadonnées sur gethub Arabert

Pour le premier algorithme, il s'agit d'un code open source sur gethub (Arabert) Figure 4 que nous avons essayé d'entraîner pour extraire les informations essentielles à partir des fichiers texte, que l'on appelle les métadonnées. Le deuxième algorithme interagit directement avec l'utilisateur, lui permettant de réaliser des recherches dans la base de données et de lui présenter les métadonnées correspondantes.

6 Résultats préliminaires

La conversion des documents numérisés en fichiers texte donne des résultats très satisfaisants Figure 5.

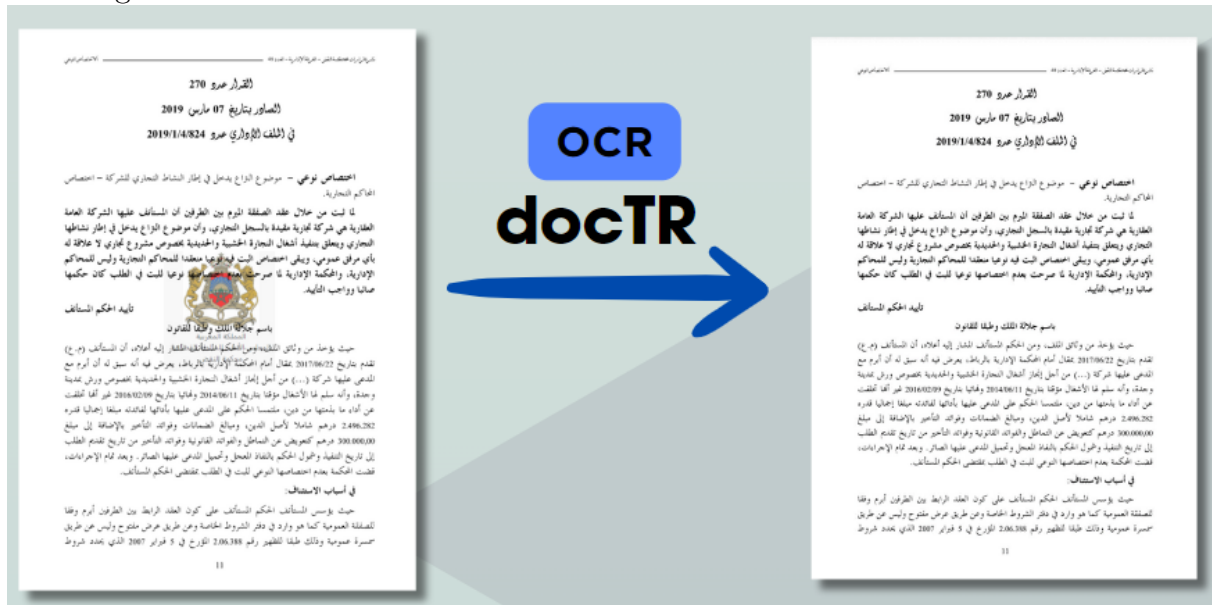


FIGURE 5 – Résultat de conversion de fichier par OCR docTR

Le temps de réponse est de 173 sec, et l'efficacité mesurée est de 93%. L'utilisation de l'algorithme de recherche développé par les étudiants de troisième année donne de bons résultats. Cependant, l'algorithme d'extraction des métadonnées nécessite davantage d'entraînement avec une base de données beaucoup plus importante pour devenir plus efficace (Figure 6). Cette nécessité d'un entraînement supplémentaire est due à l'importance de familiariser l'algorithme avec le lexique spécifique du droit. Il est essentiel que l'algorithme apprenne à reconnaître et à ne pas négliger les parties importantes dans les documents de jurisprudence, ce qui est essentiel pour assurer l'exactitude et la pertinence des informations extraites.

Résumé en arabe

وبعد المداولة طبقاً للقانون:

حيث يهدف الطلب الى الحكم بإلغاء الاذن الصادر عن السيد عامل عمالة مقاطعات الحلي الحسيني عن الشق عدد 120 وتاريخ 18/8/2020 مع ترتيب الاثار القانونية على ذلك.

وحيث دفعت شركة الخطوط الملكية المغربية بواسطة نائبها بكون الاذن بالتسريح الصادر عن السيد عامل عمالة مقاطعات الحلي الحسيني لأسباب اقتصادية لا يشكل قرارا اداريا بمفهومه القانوني متممة لاته العلة التصريح بعدم قبول الطلب.

وحيث ان المحكمة بعد الاطلاع على وثائق الملف ودراستها لكافة المعطيات وتفحصها لموضوع الطلب والقرار المطعون فيه تبين لها حقاً ان الدفع المثار من طرف السيد عامل عمالة مقاطعة عين الشق بان الاذن المطلوب منه هو ليس بقرار اداري وانما هو عمل تحضيرى منصوص عليه في القانون باعتباره عملاً تحضيرياً.

وان الجهة المشغلة هي التي يتوقف عليها الأخذ به ضمن شروط التسريح للعمال او المستخدمين وبالتالي فانه ليس بقرار اداري مستجمع الاركان حتى يمكن الطعن فيه بالإلغاء مادام هو شرط متطلب في اتخاذ قرار اداري نهائي يبقى معلقاً بيد الجهات المشغلة ولا يخص العامل وانما يخص جهة اخرى وهي التي يبقى عليها تفعيل الاذن او عدم تفعيله بخصوص قرار الاستغناء عن العمال او المستخدمين وبالتالي فان الدفع المثار بشأنه يبقى مؤسسا من هذه الناحية ويتعين اعتماده.

وحيث انه تبعا لذلك يتعين التصريح بعدم قبول الطلب.

FIGURE 6 – Exemple de résumé

7 Business Model



8 Perspectives

Les horizons de notre projet s'annoncent prometteurs, dévoilant un terrain propice à une multitude de développements à venir dans le domaine de la technologie juridique. Découvrons à présent quelques axes prometteurs qui méritent une exploration approfondie :

- **Améliorations continues grâce à l'apprentissage automatique :** En intégrant des mécanismes d'apprentissage automatique, notre système pourrait évoluer et s'adapter aux besoins spécifiques des utilisateurs. L'amélioration constante de la précision de l'extraction des données et de la génération de résumés pourrait être un axe de développement clé.
- **Collaboration avec des plateformes existantes :** En explorant des partenariats avec des plateformes juridiques existantes, notre projet pourrait être intégré de manière transparente dans les outils déjà utilisés par les professionnels du droit. Cela faciliterait l'adoption de notre solution et renforcerait son impact sur le marché.
- **Extension à d'autres domaines juridiques :** Bien que notre projet soit initialement centré sur la jurisprudence, il pourrait être étendu pour inclure d'autres domaines juridiques tels que la législation, les contrats, ou d'autres types de documents légaux. Cela élargirait la portée de notre solution et la rendrait encore plus polyvalente.

9 Conclusion

En conclusion, notre projet représente une réponse novatrice à un défi majeur dans le domaine juridique, en ciblant l'amélioration de l'accès et de l'efficacité dans l'exploitation des documents de jurisprudence. En combinant les progrès du traitement du langage naturel et de l'intelligence artificielle, nous proposons une solution automatisée qui simplifie l'extraction des données cruciales et génère des résumés concis. Cette approche vise à faciliter la vie des professionnels du droit, tels que les avocats, les juristes, et les étudiants en droit, en comblant le vide actuel dans le paysage technologique juridique.

L'opportunité sur le marché est significative, avec un large éventail d'utilisateurs potentiels, notamment les cabinets d'avocats, les avocats individuels, les chercheurs juridiques et les institutions éducatives. Notre solution apporte une valeur ajoutée en simplifiant le processus complexe de navigation et de compréhension des documents juridiques, offrant ainsi une perspective innovante dans le domaine technologique juridique.

En tant que projet pionnier, notre initiative s'inscrit dans une démarche visant à redéfinir la manière dont les professionnels du droit interagissent avec la jurisprudence. Nous sommes convaincus que notre solution aura un impact positif, non seulement en comblant les lacunes existantes, mais aussi en contribuant à l'évolution du paysage technologique juridique.