# Opération de R&D

|  |  |
| --- | --- |
| **Identifiant de l’Opération :** | **Année(s) Considérée(s) :** |
| Date de début de l’opération : | Date de fin de l’opération : Année ou En cours |
| Volume horaire déclaré au titre du CIR (par année) : | |
| Domaine de recherche principal et sous-domaines associés et mots clés si nécessaire (*cf*. [nomenclature](https://barriereconseil.sharepoint.com/sites/consulting/Documents%20partages/Forms/AllItems.aspx?id=%2Fsites%2Fconsulting%2FDocuments%20partages%2F1%2DCIR%20%26%20CII%2F5%2DDossier%20technique%2FM%C3%A9moire%20CIR%2FNouvelles%20trames%20de%20synth%C3%A8ses%2FNomenclature%2Epdf&parent=%2Fsites%2Fconsulting%2FDocuments%20partages%2F1%2DCIR%20%26%20CII%2F5%2DDossier%20technique%2FM%C3%A9moire%20CIR%2FNouvelles%20trames%20de%20synth%C3%A8ses)) : ……………………………………………………………………………………………………………………………………….. | |

## Contexte de l’opération de R&D

Voici une proposition structurée et claire pour la section "Contexte" :  
  
---  
  
### Contexte  
  
Dans un environnement où la production de contenus écrits occupe une place centrale dans la stratégie de communication des entreprises, garantir la cohérence éditoriale, la conformité aux valeurs de marque et l’efficacité opérationnelle représente un enjeu majeur, en particulier à l’échelle des grandes organisations. Les entreprises sont confrontées à la difficulté d’intégrer l’Intelligence Artificielle (IA) dans leurs processus rédactionnels tout en préservant leur identité de marque, la qualité des contenus et la conformité réglementaire (notamment en matière de sécurité, de RGPD ou de certifications comme SOC 2).  
  
Face à ces défis, Mark AI a développé une solution innovante permettant d’assurer, en temps réel, la conformité des contenus générés avec les règles éditoriales spécifiques à chaque client (orthographe, grammaire, tonalité, vocabulaire, interdictions, etc.). Cette solution s’appuie sur des modèles d’IA spécialisés, entraînés par fine-tuning à partir de grands modèles de langage (LLM) open source (tels que Mistral ou Llama) ou propriétaires (OpenAI), et adaptés à des tâches précises telles que la détection d’erreurs, la suggestion de corrections ou la génération de textes longs.  
  
L’approche de Mark AI repose sur plusieurs piliers :  
  
- \*\*Entraînement sur mesure des modèles\*\* : Chaque modèle est affiné à partir de jeux de données élaborés manuellement ou co-créés avec l’IA, garantissant un haut niveau de précision et d’adéquation avec les besoins métier.  
- \*\*Outils d’évaluation avancés\*\* : Plus de 30 évaluateurs standards et personnalisés permettent de mesurer la conformité linguistique, la précision des contenus, le respect des exigences techniques (SEO, longueur, mots-clés) et l’application des règles spécifiques à chaque client.  
- \*\*Automatisation et scalabilité\*\* : La plateforme intègre des modules permettant la génération automatique de contenus longs, la production en masse (par lot à partir de fichiers CSV) ou la création de contenus à partir de documents variés (Word, PDF, audio, vidéo), offrant ainsi un gain de temps significatif.  
- \*\*Intégration de services externes\*\* : Grâce au function calling, l’IA peut interagir avec des sources externes (recherche Google, base de connaissances client, scraping d’URL, analyse SEO, etc.) pour enrichir et fiabiliser les contenus générés.  
  
Depuis 2021, Mark AI a progressivement enrichi sa solution : d’abord centrée sur la génération de contenus marketing via des templates, elle s’est orientée vers la création de workflows pour les formats longs, puis vers la personnalisation avancée pour les entreprises (gestion fine des personas, des règles d’écriture et des bases de connaissances). En 2024, l’accent a été mis sur le développement d’outils d’évaluation internes et de modèles de détection/suggestion en temps réel, renforçant la qualité et la conformité des contenus produits.  
  
Cette approche différenciée permet à Mark AI de répondre aux exigences des grands comptes dans des secteurs sensibles (logiciel, services financiers, santé), où la maîtrise de la marque, la sécurité et la conformité sont des priorités stratégiques. La solution se distingue également par sa capacité à s’intégrer dans les processus existants des clients et à s’adapter à des besoins spécifiques, là où les solutions concurrentes restent souvent généralistes ou limitées à certains aspects de la conformité rédactionnelle.  
  
---  
  
N’hésitez pas à demander une adaptation si vous souhaitez insister sur certains aspects (historique, différenciation concurrentielle, etc.).

* Présenter **la problématique ayant justifié l’initiation d’une opération de recherche**. Notons qu’une source bibliographique présentant ou illustrant le propos sera ici utile afin de renforcer la justification de la démarche. Ladite problématique peut ici être généralisée à un domaine ou à un secteur d’activité ou bien être propre à une limite interne de l’entreprise ;
* Présenter **l’opération de R&D et justifier de la nécessité de poursuite** lorsqu’il y a une antériorité au projet ;
* **Replacer l’activité de recherche au sein de l’activité de l’entreprise** ;
* **Présenter la visée finale (et générale) de l’opération** : Il n’est pas ici nécessaire de rentrer en détail dans les objectifs poursuivis dans le cadre de la démarche énoncée. Ainsi, l’unique visée générale de la démarche sera ici présentée, ceci sans présenter chacun des objectifs faisant l’objet de l’année défendue au titre du CIR.

Outre ces éléments, il convient de rédiger cette partie d’introduction du sujet **au présent**.

*Note pour rédaction* *: En cas d’opération de R&D comportant plusieurs projets, il conviendra de les présenter sous les traits d’une opération unique en exposant un objectif et une problématique générale, adressant l’ensemble des projets.*

## Indicateurs de R&D

### Indicateurs  
  
Afin de piloter et d’objectiver la performance des travaux de R&D menés par Mark AI, un ensemble d’indicateurs a été mis en place. Ces indicateurs couvrent à la fois la qualité des modèles développés, l’efficacité des outils d’évaluation, l’impact sur la production de contenus, ainsi que l’adoption et la satisfaction client. Ils permettent de mesurer de façon quantitative et qualitative l’apport des innovations réalisées dans le cadre du Crédit d’Impôt Recherche.  
  
#### 1. Indicateurs de performance des modèles d’IA  
  
- \*\*Taux de conformité éditoriale\*\*   
 Pourcentage de contenus générés respectant l’ensemble des règles éditoriales de la marque (orthographe, grammaire, tonalité, terminologie, etc.), mesuré via les outils d’évaluation internes.  
- \*\*Précision des modèles\*\*   
 Mesure du taux d’hallucinations (informations erronées ou inventées) et du respect des sources, calculée sur des jeux de tests annotés.  
- \*\*Respect des contraintes techniques\*\*   
 Pourcentage de contenus générés respectant les contraintes de longueur, d’intégration des mots-clés SEO et de structure attendue.  
- \*\*Vitesse de génération\*\*   
 Temps moyen nécessaire à la génération d’un contenu long ou à la production en masse via le module tâches, comparé aux benchmarks du marché.  
- \*\*Taux de réussite des phases de test\*\*   
 Pourcentage de modèles fine-tunés validés dès la première itération d’évaluation.  
  
#### 2. Indicateurs de qualité des jeux de données et de l’évaluation  
  
- \*\*Volume et diversité des jeux de données\*\*   
 Nombre d’exemples utilisés par modèle (de 150 à 1000 selon la complexité), diversité des cas couverts, part de données co-créées avec l’IA.  
- \*\*Nombre et couverture des évaluateurs\*\*   
 Nombre d’outils d’évaluation standards et personnalisés développés (plus de 30 à ce jour), couverture des différents critères (linguistique, technique, éditorial, etc.).  
- \*\*Taux de validation humaine\*\*   
 Proportion d’évaluations nécessitant une validation humaine, indicateur de la robustesse des évaluateurs automatiques.  
  
#### 3. Indicateurs d’impact sur la production de contenus  
  
- \*\*Gain de temps de production\*\*   
 Temps économisé par rapport à une génération manuelle ou à l’usage de solutions concurrentes, notamment via le module tâches (génération en masse, traitement de fichiers, etc.).  
- \*\*Volume de contenus générés\*\*   
 Nombre de contenus produits par période (semaine, mois), y compris via la génération en masse à partir de fichiers CSV.  
- \*\*Taux d’adoption des fonctionnalités avancées\*\*   
 Part des utilisateurs exploitant les modules de détection/suggestion en temps réel, les outils spécialisés (scraping, analyse SERP, maillage interne, etc.).  
  
#### 4. Indicateurs de satisfaction et d’adoption client  
  
- \*\*Taux de satisfaction client\*\*   
 Mesuré via des enquêtes régulières, témoignages et retours clients (ex. : Beachcomber, Contentsquare).  
- \*\*Taux de renouvellement et d’extension\*\*   
 Pourcentage de clients renouvelant leur abonnement ou élargissant leur périmètre d’utilisation.  
- \*\*Taux de conversion commerciale\*\*   
 Indicateur de la performance commerciale (ex. : 10%+ de conversion, croissance du pipeline, augmentation de l’ACV).  
  
#### 5. Indicateurs de robustesse et de sécurité  
  
- \*\*Respect des exigences de sécurité\*\*   
 Conformité aux standards (SOC 2, RGPD), nombre d’incidents de sécurité ou de non-conformité détectés.  
- \*\*Disponibilité et performance de la plateforme\*\*   
 Taux de disponibilité des services, temps de réponse moyen des modèles en production.  
  
---  
  
L’ensemble de ces indicateurs constitue un tableau de bord de suivi permettant d’ajuster en continu les efforts de R&D, d’objectiver les avancées réalisées, et de démontrer la valeur ajoutée des travaux éligibles au Crédit d’Impôt Recherche.

* Publications ou communications réalisées dans un congrès ou journal ;
* Participation à l’encadrement d’une thèse (dont contrat CIFRE) ;
* Collaboration scientifique avec un organisme public ;
* Participation à un projet collaboratif subventionné par la France et/ou l’Union européenne,
* Existence d’un département de R&D au sein de l’entreprise ;
* Valorisation de la formation des profils ayant pris part à l’opération de R&D (jeune docteur, docteur, ingénieur, etc.).

Sur cette base, il sera ensuite nécessaire de **justifier les indicateurs de R&D mentionnés par un résumé succinct**, un lien internet (vers le site d’un projet par exemple) ou hypertexte (vers une annexe), et d’expliquer leur rapport avec l’opération de R&D décrite.

Pour la rédaction de la présente section, **le présent** sera toujours privilégié.

## Objet de l’opération de R&D

Section : Objectifs  
  
L'objectif principal du projet Mark AI est de développer une plateforme innovante d’assistance à la rédaction basée sur l’intelligence artificielle, permettant aux entreprises de garantir la conformité éditoriale, la cohérence de la marque et la qualité des contenus produits à grande échelle. Pour répondre aux enjeux spécifiques des grandes organisations (protection de la réputation, conformité réglementaire, efficacité opérationnelle), la solution vise à offrir des outils avancés, personnalisables et intégrés, reposant sur des modèles d’IA propriétaires et des systèmes d’évaluation robustes.  
  
Les objectifs scientifiques, techniques et opérationnels du projet sont les suivants :  
  
1. \*\*Développer des modèles d’IA spécialisés et performants\*\*   
 - Concevoir et entraîner, par fine-tuning, des modèles de langage adaptés à des tâches précises de vérification et de suggestion éditoriale (orthographe, grammaire, tonalité, application des règles de marque, détection de formulations IA, etc.).  
 - Garantir la précision, la fiabilité et la rapidité des générations en s’appuyant sur des modèles open source et propriétaires, optimisés pour l’intégration en temps réel dans l’éditeur de texte.  
  
2. \*\*Automatiser et fiabiliser la conformité éditoriale à l’échelle de l’entreprise\*\*   
 - Permettre la détection automatique des écarts par rapport aux lignes éditoriales et aux règles de marque (glossaires, terminologies, tonalités, personas, etc.), et proposer des corrections ou suggestions en temps réel.  
 - Assurer la conformité linguistique, technique (SEO, longueur, structure) et informationnelle (sources, absence d’hallucinations) des contenus générés.  
  
3. \*\*Industrialiser la production de contenus personnalisés et de qualité\*\*   
 - Développer des modules permettant la génération de contenus longs ou en masse (par lot ou à partir de fichiers structurés), tout en maintenant un haut niveau de personnalisation et de respect des consignes éditoriales propres à chaque client.  
 - Intégrer des outils d’enrichissement automatique des contenus (recherche documentaire, scrapping, analyse concurrentielle, maillage interne, etc.) pour garantir la pertinence et l’actualité des informations.  
  
4. \*\*Mettre en place un système d’évaluation de la qualité et de suivi continu\*\*   
 - Concevoir une suite d’outils d’évaluation programmatique, hybride (code/LLM) et humaine, permettant de mesurer objectivement la qualité des générations (conformité, précision, respect des instructions).  
 - Assurer un contrôle qualité systématique avant la mise en production de nouveaux modèles ou de modifications, et un suivi continu des performances en exploitation.  
  
5. \*\*Répondre aux exigences de sécurité, de confidentialité et de scalabilité des entreprises\*\*   
 - Garantir l’hébergement sécurisé des modèles et des données (self-hosted, conformité SOC 2 et RGPD).  
 - Offrir une solution modulable, adaptée aux besoins de différents secteurs (logiciel, finance, santé), et capable de s’intégrer dans les processus de production de contenu à grande échelle.  
  
6. \*\*Se différencier durablement de la concurrence internationale et nationale\*\*   
 - Proposer une solution combinant personnalisation avancée, rapidité, précision, sécurité et intégration native des règles de marque, là où les solutions concurrentes se limitent à des fonctionnalités génériques ou à la correction linguistique de base.  
  
En résumé, le projet Mark AI a pour ambition d’apporter une réponse technologique de pointe aux enjeux de conformité, de qualité et d’efficacité de la communication écrite en entreprise, en s’appuyant sur l’innovation en intelligence artificielle appliquée à la rédaction et à la gestion de la marque.

En outre, **pour chaque objectif énoncé, un état de l’art sera réalisé** afin de mettre en avant le caractère inédit de la visée finale au regard des connaissances existantes. En ce sens, il conviendra de s’appuyer sur des publications scientifiques ou techniques (Thèses, rapports de recherches, articles scientifiques, *etc*.). Concernant le point particulier des brevets, ces derniers peuvent être présentés néanmoins, il est préférable de privilégier les documents présentés précédemment.

Les publications doivent être annoncées entre crochet [NOM, **ANNEE**], puis commentées. Dans ce cadre, **la description des publications devra se faire au présent lorsque la mise en regard avec les travaux présenter sera au passé**. Citer entre guillemets et en italique les passages pertinents. En fin de document, donner la référence complète.

Exemple : D’après les travaux de K-F. Chow et *al*., leur nouvelle technique de détection de l’hybridation des oligonucléotides permet de « *contrôler le potentiel interfacial de multiples électrodes bipolaires bien définies, à l'aide de seulement deux fils* » [CHOW, **2008**]. Néanmoins, aucune donnée quantitative n’est ici décrite. Par conséquent, nous ne pouvions en rien nous reposer sur ces éléments.

Chaque publication doit être :

* **Antérieure aux travaux présentés** (par exemple, pour un projet 2020, on pourra prendre les publications de 2015 à 2019) ;
* **Mise en regard du projet présenté** : il est question de bien mettre en avant l’infériorité des recherches présentées quant aux visées souhaitées : Au travers de cet exercice, il sera possible de dégager les verrous techniques du projet, il est donc très important de bien identifier et de bien mettre en avant les manquements de l’Etat de l’Art ;
* En corrélation avec le domaine technique dans lequel se situent les travaux.

*Note pour rédaction : En cas de poursuite de travaux, il est possible de reprendre l’état de l’art réalisé l’année précédente. Néanmoins, il convient de vérifier sa pertinence au regard des nouvelles problématiques du projet (il est ici possible de ne reprendre qu’une partie de l’EA). De plus, s’il apparaît toujours en accord avec ces dernières, une mise à jour sera nécessaire (ajout de plusieurs publication) tous les ans.*

**Chaque état de l’art, lié à un objectif, devra ensuite se conclure sur le verrou correspondant**. Pour ce faire, il sera uniquement nécessaire, après une rapide phrase d’introduction, de formuler le verrou.

Cette section verrou est capitale dans le cadre de la valorisation d’un projet de recherche. En effet, elle synthétise toute la dimension R&D d’un projet. Par conséquent, il est très important de la construire avec concision mais précision.

Pour cela, il est possible de présenter les verrous sous forme de question. Il sera ici possible d’adjoindre à chaque verrou une rapide description afin de préciser ce dernier. En matière de rédaction pure, **le verrou sera présenté au présent.**

D’autre part, il faut bien veiller à **ne pas multiplier les verrous techniques**. En effet, un unique verrou peut tout à fait être suffisant.

En définitive, la structure du document sera donc la suivante : **N objectifs/N parties de l’EA/N verrous/N parties de travaux.**

*Point d’attention : Il est important de bien différencier la problématique de l’objectif : En effet, l’objectif est la visée finale à atteindre, la problématique est le problème technique empêchant d’atteindre cette dernière. Il faut donc ici bien veiller à ne pas transformer l’objectif, précédemment énoncé, en question.*

## Description de la démarche suivie et des travaux réalisés

### Section : Travaux Les travaux de R&D menés dans le cadre du projet Mark AI s’articulent autour de plusieurs axes majeurs, visant à développer une solution avancée d’assistance à la rédaction conforme à l’identité de marque et aux exigences éditoriales des entreprises. Ces travaux s’inscrivent dans une démarche d’innovation continue, couvrant l’entraînement de modèles d’IA, la création de jeux de données propriétaires, la mise en place de systèmes d’évaluation automatisés, ainsi que l’intégration de fonctionnalités spécialisées et l’optimisation des workflows de génération de contenus. 1. \*\*Développement et entraînement de modèles spécialisés\*\* L’un des axes principaux des travaux porte sur l’entraînement de modèles de langage spécialisés via la méthode du fine-tuning. À partir de modèles de base open source (Mistral, Llama 3B) ou propriétaires (OpenAI GPT-4o, GPT-4o-mini), l’équipe développe des modèles adaptés à des tâches précises telles que la détection en temps réel des écarts de conformité (orthographe, grammaire, tonalité, respect des règles éditoriales, détection de formulations IA, etc.) et la suggestion de corrections directement intégrées à l’éditeur de texte. L’entraînement de ces modèles s’appuie sur des jeux de données soigneusement construits, permettant d’obtenir un haut niveau de précision, de rapidité et de fiabilité dans les résultats. 2. \*\*Création et enrichissement de jeux de données propriétaires\*\* La qualité des modèles repose sur la constitution de jeux de données adaptés à chaque cas d’usage. Ces jeux de données sont élaborés principalement de manière manuelle, afin de garantir la pertinence et la représentativité des exemples. Selon la complexité des tâches, le volume d’exemples varie de 150 à 1000, avec un recours ponctuel à la co-création assistée par IA. Cette démarche permet d’adresser aussi bien des tâches simples de conformité linguistique que des tâches plus créatives nécessitant une adaptation fine au contexte de chaque client. 3. \*\*Développement d’un système d’évaluation automatisé et hybride\*\* Afin de garantir la qualité des contenus générés et le respect des lignes éditoriales, Mark AI a développé une suite d’outils d’évaluation sophistiqués. Ces outils combinent plusieurs méthodes : - \*\*Évaluation programmatique\*\* (Python) pour les critères objectifs et quantifiables, - \*\*Analyse hybride code/LLM\*\* pour les critères complexes (tonalité, style, application des personas), - \*\*Validation humaine\*\* pour les cas spécifiques ou les besoins clients particuliers. Chaque génération de contenu peut ainsi être évaluée selon différents critères (conformité linguistique, précision, conformité technique, etc.), avec des mesures numériques, des pourcentages et des validations binaires. Plus de 30 évaluateurs standards sont disponibles, auxquels s’ajoutent des évaluateurs personnalisés selon les besoins des clients. 4. \*\*Conception et intégration de modules et fonctionnalités avancées\*\* Les travaux ont également porté sur la création de modules innovants, tels que : - \*\*Le module “tâches”\*\* permettant la génération de contenus longs en arrière-plan (gain de temps significatif), la génération massive de contenus à partir de fichiers CSV, ou la rédaction à partir de fichiers variés (Word, PDF, audio, vidéo). - \*\*L’intégration de fonctionnalités de function calling\*\*, permettant à l’IA d’interagir dynamiquement avec des services externes (recherche Google, interrogation de bases de connaissances, scraping d’URL, analyse SERP, calcul d’itinéraires, etc.), afin d’enrichir les contenus générés avec des informations fiables et actualisées. - \*\*L’analyse sémantique et le maillage interne\*\*, pour optimiser le référencement naturel et la cohérence des contenus produits. 5. \*\*Optimisation des workflows et personnalisation avancée\*\* L’équipe a conçu des workflows permettant la création de contenus longs et structurés (articles de blog, descriptions produits, landing pages, etc.), adaptés aux contraintes SEO et à la personnalisation de la marque (personas, tonalité, terminologie, règles d’écriture). Ces workflows sont évolutifs et personnalisables, offrant une grande flexibilité pour répondre aux besoins spécifiques des clients, notamment dans les secteurs exigeants tels que la finance, la santé ou le software. 6. \*\*Veille technologique et comparative\*\* Enfin, une veille active sur les solutions concurrentes internationales et françaises est réalisée, afin d’identifier les axes de différenciation et d’anticiper les évolutions du marché. Cette veille guide les choix technologiques et oriente les priorités de développement. \*\*Conclusion\*\* L’ensemble de ces travaux contribue à la création d’une plateforme robuste, innovante et hautement personnalisable, permettant aux entreprises d’assurer la conformité, la qualité et la cohérence de leurs communications écrites à grande échelle, tout en intégrant les dernières avancées en intelligence artificielle appliquée au traitement du langage naturel.

Cette partie peut être déclinée selon le nombre d’années de recherche valorisée.

Tout l’enjeu est ici celui de **présenter la démarche itérative avec précision mais concision**. **Chaque partie correspondra ainsi aux travaux relatifs à la levée des problématiques précédemment présentées (1 sous-section = une problématique).**

En matière de présentation, il conviendra de mettre en lumière les différentes hypothèses de recherche énoncées et les travaux afférents, ce en se concentrant sur la résolution des problématiques techniques rencontrées.

Concernant les règles de rédaction, il sera important de :

* **Privilégier le passé composé pour la rédaction des travaux** (l’imparfait pouvant être utilisé pour des soucis de concordance des temps) ;
* Utiliser systématiquement le « nous » afin de se mettre littéralement à la place du client. Le « on » ne devra jamais être utilisé ;
* Ne jamais abréger les mots. On n’utilisera uniquement les abréviations consacrées, qui devront apparaître en italique, comme *etc*. ou *al*. par exemple ;
* Faire apparaître chaque mot en anglais ou latin (ou dans une autre langue) en italique ;
* Expliciter chaque mot technique ou spécifique en note de bas de page ;
* Garder un certain niveau de langage (aucune familiarité de langage) ;
* Faire un renvoi pour chaque figure présentée soit en indiquant : (cf. Figure 1), soit en intégrant le renvoi à une phrase : comme présenté en figure 1 (ici, le mot figure comportera un f minuscule) ;
* Faire un renvoi aux annexes présentées selon le modèle : cf. annexe 1.1 : « *Titre de l’annexe* », cf. annexe 1.2 : « *Titre de l’annexe* », cf. annexe 1. 3 : « *Titre de l’annexe* ».
* En cas d’information manquante, ne pas hésiter à formaliser, directement dans le corps du texte, une question au client.

## Ressources Humaines

|  |  |  |
| --- | --- | --- |
| Personnel R&D | Heures R&D | Rôle au sein de l’opération de R&D |
|  |  |  |
|  |  |  |
|  |  |  |

## Contribution scientifique, technique ou technologique

Bien sûr, voici une proposition structurée et claire pour la section "Contribution", adaptée au contexte du Crédit d’Impôt Recherche (CIR) :  
  
---  
  
## Contribution  
  
Les travaux de R&D menés par Mark AI ont permis de réaliser des avancées significatives dans le domaine de la génération et de la conformité éditoriale assistées par intelligence artificielle, avec une approche centrée sur la personnalisation, la qualité et la robustesse des contenus générés pour les entreprises. Les contributions majeures se déclinent selon plusieurs axes :  
  
### 1. Développement de modèles IA spécialisés et innovants  
  
Mark AI a conçu et entraîné plusieurs dizaines de modèles d’intelligence artificielle spécialisés, reposant sur la technique du fine-tuning de grands modèles de langage (LLM) open source (Mistral, Llama) et propriétaires (OpenAI). Ces modèles sont adaptés à des tâches très spécifiques, telles que :  
  
- La détection en temps réel des non-conformités avec les règles éditoriales propres à chaque marque (orthographe, grammaire, tonalité, termes interdits, langage IA, etc.).  
- La suggestion automatisée de corrections et d’améliorations directement intégrées à l’éditeur de texte.  
- La génération de contenus longs et complexes, en arrière-plan, permettant de traiter des volumes importants (jusqu’à plusieurs milliers de contenus générés en un clic à partir d’un fichier CSV ou de différents formats d’entrée : Word, PDF, audio, vidéo).  
  
### 2. Création et structuration de jeux de données propriétaires  
  
La qualité des modèles repose sur l’élaboration de jeux de données propriétaires, adaptés à chaque cas d’usage. Mark AI a mis en place des processus rigoureux de constitution de ces jeux de données :  
  
- Annotation manuelle de plusieurs centaines à milliers d’exemples pour chaque tâche (conformité linguistique, créativité, respect de la tonalité, etc.).  
- Co-création assistée par IA pour accélérer la génération d’exemples tout en maintenant un haut niveau de qualité et de représentativité.  
- Adaptation systématique des jeux de données aux besoins spécifiques de chaque client ou secteur d’activité.  
  
### 3. Développement d’un système d’évaluation avancé  
  
Pour garantir la performance et la fiabilité des modèles, Mark AI a développé une suite complète d’outils d’évaluation, permettant de mesurer objectivement la qualité des contenus générés et leur conformité aux exigences éditoriales :  
  
- Évaluation programmatique (Python) pour les critères simples et quantifiables.  
- Analyse hybride (code/LLM) pour les critères complexes tels que la tonalité ou le style rédactionnel.  
- Validation humaine pour les cas nécessitant une expertise spécifique.  
- Mise en place de plus de 30 évaluateurs standards, ainsi que d’évaluateurs personnalisés pour répondre aux besoins particuliers de certains clients.  
  
### 4. Intégration d’outils spécialisés et d’interfaces IA avancées  
  
Mark AI a enrichi sa plateforme avec des outils permettant à l’IA d’interagir dynamiquement avec des services externes, renforçant la pertinence et la fiabilité des contenus générés :  
  
- Recherche Google automatisée, scraping de pages web, et synthèse d’informations pour enrichir les contenus.  
- Interrogation de bases de connaissances propriétaires des clients pour garantir la véracité des réponses.  
- Analyse sémantique pour le maillage interne et l’intégration de liens pertinents.  
- Outils de calcul d’itinéraires, d’analyse SERP et de veille concurrentielle.  
  
### 5. Apport à l’état de l’art et différenciation sur le marché  
  
Les innovations de Mark AI se distinguent par :  
  
- L’adaptation fine des modèles aux besoins de chaque marque, garantissant une conformité éditoriale inégalée.  
- L’automatisation de la génération de contenus à grande échelle, avec des gains de productivité substantiels pour les équipes de communication.  
- La mise en œuvre de solutions auto-hébergées, robustes et sécurisées, adaptées aux exigences des grandes entreprises (conformité SOC 2, RGPD, etc.).  
- Une approche orientée “qualité continue”, avec un suivi systématique des performances des modèles et une amélioration itérative basée sur les retours clients et les résultats des évaluations.  
  
---  
  
\*\*En synthèse\*\*, les travaux réalisés par Mark AI contribuent directement à l’avancement des technologies d’IA appliquées à la rédaction, à la conformité éditoriale et à la personnalisation de contenus, tout en répondant aux enjeux de sécurité, de scalabilité et d’intégration propres aux environnements d’entreprise. Ces contributions s’inscrivent pleinement dans le périmètre du Crédit d’Impôt Recherche, tant par leur caractère innovant que par la complexité scientifique et technique des problématiques traitées.

En outre, un projet de recherche se doit de pouvoir dépasser les frontières de son propre contexte de développement et ainsi être en mesure d’être externalisé à d’autres cas. Dans ce cadre, il sera nécessaire de démontrer les apports du projet au regard du domaine. Ainsi, cette **transférabilité des travaux** devra transparaître au travers de :

* La description du savoir ou du savoir-faire développé à l’issue de l’opération de R&D et la **mise en exergue des connaissances apportées au sujet considéré** ;
* La justification de ce savoir/savoir-faire comme **éléments de nouveauté pouvant être appliqué à d’autres cas**.

## Partenariat scientifique et recherche confiée

### Partenariat  
  
Dans le cadre de ses travaux de R&D, Mark AI s’appuie sur une politique de partenariat stratégique, tant sur le plan technologique que sur le plan métier, afin d’accélérer l’innovation, d’assurer la robustesse de ses solutions et de garantir leur adéquation avec les besoins réels des utilisateurs finaux.  
  
#### 1. Partenariats technologiques  
  
Mark AI collabore avec des acteurs majeurs de l’écosystème de l’intelligence artificielle et du cloud pour offrir des solutions à la pointe de la technologie. Ces partenariats permettent notamment :  
  
- \*\*Accès à des modèles de base avancés\*\* : Mark AI utilise et adapte des modèles open source reconnus (Mistral, Llama 3B) ainsi que des modèles propriétaires de partenaires leaders (OpenAI : GPT-4o-mini, GPT-4o), garantissant ainsi un haut niveau de performance, de sécurité et de flexibilité dans le développement de ses propres modèles fine-tunés.  
- \*\*Intégration de services externes\*\* : Grâce à des collaborations avec des fournisseurs de services spécialisés (API de recherche Google, outils de scraping, solutions d’analyse SERP, etc.), Mark AI enrichit ses fonctionnalités IA et assure une interopérabilité optimale avec les outils métiers de ses clients.  
- \*\*Développement d’outils d’évaluation\*\* : Les partenariats technologiques favorisent le développement d’outils d’évaluation avancés, permettant de mesurer la qualité des générations IA selon des critères linguistiques, techniques et métiers.  
  
#### 2. Partenariats métiers et co-innovation  
  
Mark AI adopte une démarche de co-innovation avec ses clients et partenaires métiers, en particulier dans les secteurs du logiciel, des services financiers et de la santé. Cette approche collaborative se traduit par :  
  
- \*\*Co-création de jeux de données\*\* : Mark AI travaille en étroite collaboration avec ses clients pour élaborer des jeux de données représentatifs de leurs besoins spécifiques (règles éditoriales, glossaires, tonalités de marque, etc.), garantissant ainsi la pertinence et la qualité des modèles développés.  
- \*\*Développement de fonctionnalités sur-mesure\*\* : En partenariat avec des entreprises clientes, Mark AI conçoit et intègre des fonctionnalités personnalisées (évaluateurs spécifiques, modules d’automatisation, intégration de bases de connaissances propriétaires), assurant une adaptation fine aux enjeux métiers.  
- \*\*Tests et validation en conditions réelles\*\* : Les solutions développées sont systématiquement testées et validées en collaboration avec les utilisateurs finaux, permettant d’itérer rapidement et d’optimiser la performance des outils.  
  
#### 3. Veille concurrentielle et benchmark international  
  
Mark AI entretient des relations suivies avec des acteurs internationaux et français du secteur de la génération de contenus assistée par IA (Writer.com, Jasper.ai, Copy.ai, Copysmith, Semji, MerciApp, etc.), ce qui lui permet :  
  
- \*\*D’identifier les meilleures pratiques\*\* et innovations du marché,  
- \*\*De benchmarker ses solutions\*\* afin d’assurer un positionnement différenciant, notamment sur la conformité éditoriale, la personnalisation avancée et la sécurité des données,  
- \*\*D’anticiper les évolutions technologiques\*\* et réglementaires pour rester à la pointe de l’innovation.  
  
#### 4. Collaboration avec l’écosystème académique et institutionnel  
  
Enfin, Mark AI s’inscrit dans une dynamique d’ouverture vers l’écosystème académique et institutionnel, notamment via :  
  
- \*\*La participation à des groupes de travail sectoriels\*\* sur l’IA et la conformité éditoriale,  
- \*\*La veille scientifique\*\* pour intégrer les avancées récentes en NLP et en évaluation automatique,  
- \*\*La contribution à des projets collaboratifs\*\* visant à améliorer la qualité, la sécurité et l’éthique des solutions d’IA générative.  
  
---  
  
Grâce à cette politique de partenariat structurée, Mark AI renforce la qualité, la pertinence et la compétitivité de ses innovations, tout en favorisant l’émergence de solutions robustes et adaptées aux enjeux des entreprises clientes.

* Dans le cas d’un partenariat, expliquer le rôle de l’entreprise dans le consortium de R&D ;
* Pour les travaux de sous-traitance, préciser lesquels en décrivant la contribution du sous-traitant.

## Références bibliographiques

Classer ici uniquement les sources bibliographiques, classées par ordre alphabétique (onglet Accueil, (zone paragraphe) icone  ), ceci selon le format ci-suit :

[NOM, ANNEE] NOM, P., Titre, Journal, ANNEE, vol. p.

Ex : CHOW, KF. et al. Wireless electrochemical DNA microarray sensor, JACS, 2008, vol. 130, p. 7544

Ou issue de Google Scholar, fonction « citer », c/c de la norme ISO 690 :

[COSTENTIN, 2013] COSTENTIN, C. et al. Catalysis of the electrochemical reduction of carbon dioxide. Chemical Society Reviews, 2013, vol. 42, no 6, p. 2423-2436

HILL, H. et al. Electrochemical assay for nucleic acids and nucleic acid probes. U.S. Patent No 4,840,893, 20 juin 1989