**CREDIT D’IMPOT EN FAVEUR DE LA RECHERCHE**

**CREDIT D’IMPOT EN FAVEUR DE LA RECHERCHE**

EXERCICE

20XX

**CLIENT**

**LOGO**

Dossier Justificatif Confidentiel

**CREDIT D’IMPOT EN FAVEUR DE LA RECHERCHE**

SOMMAIRE

[PRESENTATION DE L’ENTREPRISE 2](#_Toc88587594)

[1. L’entreprise 3](#_Toc88587595)

[2. Gestion de la recherche 4](#_Toc88587596)

[3. Situation vis-à-vis du Crédit d’Impôt Recherche 5](#_Toc88587597)

[ELEMENTS DE VALORISATION DU CREDIT D’IMPOT RECHERCHE 20XX 6](#_Toc88587598)

[1. Ventilation nominative par pro 7](#_Toc88587599)

[2. Dépenses de personnel 8](#_Toc88587600)

[3. Dotations aux amortissements des immobilisations affectées à la recherche 9](#_Toc88587601)

[4. Montant encaissé des subventions publiques 10](#_Toc88587602)

[5. Etat récapitulatif 11](#_Toc88587603)

[ELEMENTS DE VALORISATION DU CREDIT D’IMPOT RECHERCHE 20XX 12](#_Toc88587604)

[1. Opération de R&D 13](#_Toc88587605)

[1.1. Contexte de l’opération de R&D 13](#_Toc88587606)

[1.2. Indicateurs de R&D 13](#_Toc88587607)

[1.3. Objet de l’opération de R&D 14](#_Toc88587608)

[1.4. Description de la démarche suivie et des travaux réalisés 15](#_Toc88587609)

[1.5. Ressources Humaines 15](#_Toc88587611)

[1.6. Contribution scientifique, technique ou technologique 15](#_Toc88587612)

[1.7. Partenariat scientifique et recherche confiée 16](#_Toc88587613)

[1.8. Références bibliographiques 16](#_Toc88587614)

[ANNEXES 17](#_Toc88587615)

[ANNEXES X.X 18](#_Toc88587616)

PRESENTATION DE L’ENTREPRISE

# L’entreprise

## Présentation de l’entreprise  
  
### 1. Présentation générale de l’entreprise  
  
\*\*Raison sociale\*\* : RAG (Raison sociale à confirmer par le client)   
\*\*Secteur d’activité\*\* : Technologies de l’Information, Intelligence Artificielle, Solutions logicielles pour le conseil et la gestion documentaire automatisée   
\*\*Forme juridique\*\* : [À compléter par le client]   
\*\*Localisation\*\* : [À compléter par le client – ex. Paris, France]   
\*\*Date de création\*\* : [À compléter par le client – ex. 2022]   
\*\*Taille de l’entreprise\*\* : PME innovante (effectif à préciser)   
\*\*Marché cible\*\* : Cabinets de conseil, sociétés de services, départements de veille et d’analyse, entreprises nécessitant une automatisation avancée de la production documentaire  
  
### 2. Principales activités et domaines d’expertise  
  
RAG se positionne comme un acteur innovant dans le développement de solutions logicielles intégrant l’intelligence artificielle, avec une spécialisation dans :  
  
- \*\*Recherche documentaire automatisée et multi-source\*\* : Extraction, indexation et analyse de contenus issus du web, de documents PDF, d’images et de bases de données structurées.  
- \*\*Génération automatisée de livrables\*\* : Production rapide et personnalisée de rapports, présentations et newsletters à partir de modèles préformatés et de modules IA avancés.  
- \*\*Agents conversationnels et interactifs\*\* : Développement d’agents intelligents capables de comprendre, reformuler et répondre à des questions métiers complexes.  
- \*\*Optimisation des processus métiers\*\* : Automatisation de la veille, de l’analyse et de la rédaction pour les consultants, avec une interface conviviale et accessible sans expertise technique.  
- \*\*Intégration de technologies de pointe\*\* : Utilisation des dernières avancées en NLP, Retrieval-Augmented Generation (RAG), agents multi-modaux, et architectures cloud (Microsoft Azure).  
  
### 3. Organigramme simplifié et structure R&D  
  
L’entreprise accorde une place centrale à la Recherche et Développement, moteur de son innovation.   
\*\*Structure simplifiée :\*\*  
  
- \*\*Direction Générale\*\*  
 - \*\*Direction Technique (CTO)\*\*  
 - \*\*Équipe R&D / IA\*\*  
 - Cheffe de projet R&D / Data Scientist (ex. Chaima Saadi)  
 - Ingénieurs IA et Data  
 - Développeurs Full Stack  
 - Experts en NLP et systèmes multi-agents  
 - \*\*Équipe Produit\*\*  
 - Product Owner  
 - UX/UI Designer  
 - \*\*Direction Commerciale & Conseil\*\*  
 - Consultants métiers  
 - Support client  
  
La R&D représente le cœur de l’activité, avec une équipe pluridisciplinaire dédiée à l’expérimentation, au prototypage et à l’industrialisation des solutions innovantes.  
  
### 4. Historique et évolutions majeures  
  
- \*\*2022\*\* (date à confirmer) : Création de RAG, avec pour ambition de révolutionner la production documentaire dans le secteur du conseil grâce à l’IA.  
- \*\*2023\*\* : Constitution de l’équipe R&D et lancement des premiers prototypes d’agents conversationnels spécialisés.  
- \*\*2024\*\* : Développement et déploiement de la plateforme AiQo Search Gen Ai, intégrant des modules avancés de recherche documentaire, de génération de livrables multi-formats et d’agents intelligents.   
 - Intégration des dernières innovations en RAG, agentic workflows et hybrid retrieval.  
 - Début de la commercialisation auprès de cabinets de conseil et d’entreprises partenaires.  
- \*\*2025\*\* (prévisionnel) : Extension des fonctionnalités (memory graph, fine-tuning sectoriel), renforcement des partenariats technologiques.  
  
### 5. Chiffres clés (à compléter par le client)  
  
- \*\*Effectif total\*\* : [À préciser – ex. 10 collaborateurs, dont 60% en R&D]  
- \*\*Chiffre d’affaires 2023\*\* : [À compléter]  
- \*\*Part du chiffre d’affaires consacrée à la R&D\*\* : [À compléter – ex. 35%]  
- \*\*Nombre de projets R&D actifs\*\* : [À compléter]  
- \*\*Clients principaux\*\* : Cabinets de conseil, sociétés de services, entreprises innovantes  
  
### 6. Mise en valeur des activités R&D et innovation  
  
RAG consacre une part significative de ses ressources humaines et financières à la R&D, avec une équipe dédiée à l’expérimentation de technologies de rupture (RAG, agents multi-modaux, hybrid retrieval, memory graphs, etc.).   
L’entreprise se distingue par :  
  
- L’intégration rapide des avancées scientifiques issues de l’état de l’art (Self-RAG, agentic workflows, graph-based retrieval…)  
- La capacité à transformer ces innovations en solutions opérationnelles, robustes et adaptées aux besoins métiers des consultants.  
- Un engagement fort dans la veille technologique, la publication et la collaboration avec des acteurs académiques et industriels.  
  
---  
  
\*\*NB\*\* : Certaines informations (effectifs, CA, pourcentage R&D, forme juridique, localisation précise) sont à compléter par le client pour finalisation du dossier CIR.

# Gestion de la recherche

# Gestion de la recherche  
  
## Organisation de la Recherche  
  
La société RAG a structuré son activité de Recherche et Développement (R&D) autour d’une organisation agile et collaborative, adaptée à la nature innovante de ses projets. La direction de la recherche est assurée par un responsable R&D, garant de la cohérence scientifique et technique des travaux menés. Ce responsable coordonne une équipe pluridisciplinaire composée d’ingénieurs spécialisés en intelligence artificielle, data science, développement logiciel et gestion documentaire avancée.   
La coordination des projets s’effectue via des réunions hebdomadaires, des outils de gestion de projet (type Jira/Confluence) et des points d’avancement réguliers avec la direction technique. Un comité scientifique interne, composé de référents techniques et de responsables de pôles, valide les orientations stratégiques et les choix technologiques majeurs.  
  
\*À compléter par le client : organigramme détaillé, noms et fonctions des responsables, fréquence des comités, etc.\*  
  
---  
  
## Expertise des équipes de Recherche  
  
Les équipes de R&D de RAG sont constituées de profils hautement qualifiés, disposant d’une double compétence en ingénierie informatique et en intelligence artificielle appliquée. Elles regroupent :  
  
- \*\*Ingénieurs en génie informatique\*\* (niveau master ou ingénieur, spécialisation data science, IA, traitement automatique du langage naturel)  
- \*\*Développeurs full stack\*\* maîtrisant les environnements Python, Java, JavaScript, NodeJS, ReactJS, SQL/NoSQL  
- \*\*Experts en IA générative\*\* (LLM, Transformers, NLP avancé, modèles RAG, hybrid retrievers)  
- \*\*Spécialistes en extraction et traitement de données multimodales\*\* (textes, images, tableaux, PDF)  
- \*\*Compétences complémentaires\*\* : gestion de projets innovants, veille technologique, expérimentation sur plateformes cloud (Azure, Power BI), maîtrise des méthodologies agiles  
  
Les membres de l’équipe bénéficient d’une formation continue sur les dernières avancées en IA, RAG, et automatisation documentaire, et participent régulièrement à des conférences et séminaires spécialisés.  
  
\*À compléter par le client : CV synthétiques, diplômes, certifications, formations suivies, publications éventuelles.\*  
  
---  
  
## Typologie des projets de Recherche  
  
RAG conduit une R&D à la croisée de la recherche appliquée et expérimentale, en réponse aux besoins métiers du secteur du conseil :  
  
- \*\*Recherche appliquée\*\* : développement de nouveaux agents conversationnels, intégration de frameworks innovants (AFlow, LangGraph, Self-RAG, TCAF), amélioration des processus d’automatisation documentaire.  
- \*\*Recherche expérimentale\*\* : validation de nouveaux modèles de RAG hybrides, expérimentation de méthodes avancées de chunking, reranking et graph-based retrieval (HippoRAG), tests sur des corpus multi-formats et multi-langues.  
- \*\*Recherche fondamentale\*\* (le cas échéant) : veille sur les architectures de LLM, exploration de nouveaux paradigmes d’interopérabilité agentique.  
  
Les projets sont menés selon des cycles itératifs (proof of concept, prototypage, expérimentation, industrialisation), avec une documentation rigoureuse des verrous scientifiques et des résultats obtenus.  
  
---  
  
## Critères de sélection des projets CIR  
  
La sélection des projets éligibles au CIR repose sur une évaluation rigoureuse des critères suivants :  
  
- \*\*Caractère innovant et incertain\*\* : existence d’un verrou scientifique ou technique clairement identifié (ex. : articulation optimale entre agents conversationnels, modèles adaptatifs et dispositifs de recherche documentaire multi-source).  
- \*\*Méthodologie scientifique\*\* : formulation d’hypothèses, état de l’art, expérimentation, analyse comparative, validation des résultats.  
- \*\*Apport original\*\* : développement de solutions non disponibles sur le marché ou significativement améliorées par rapport à l’existant.  
- \*\*Documentation\*\* : traçabilité des travaux, rapports d’expérimentation, relevés de tests, suivi des versions.  
- \*\*Impact sur l’activité\*\* : contribution directe à l’amélioration des performances, de la qualité ou de la compétitivité de l’entreprise.  
  
Un comité interne d’évaluation (composé du responsable R&D, du référent CIR et de la direction technique) valide l’éligibilité des projets sur la base de ces critères.  
  
\*À compléter par le client : grille d’évaluation interne, procédures formalisées, exemples de projets retenus.\*  
  
---  
  
## Matériels et infrastructures utilisés pour la R&D  
  
Afin de soutenir ses travaux de recherche, RAG s’appuie sur une infrastructure technique robuste et évolutive :  
  
- \*\*Environnements cloud\*\* (Microsoft Azure) : déploiement de modèles LLM, orchestration de micro-services, stockage sécurisé des données (private VNet, RBAC Entra ID)  
- \*\*Bases de données vectorielles et relationnelles\*\* : pour l’indexation, la recherche sémantique et la gestion documentaire  
- \*\*Outils de scraping et de traitement de données\*\* : BeautifulSoup, frameworks de web scraping, pipelines d’extraction PDF et images  
- \*\*Environnements de développement\*\* : Python, Java, JavaScript, TensorFlow, Pandas, Numpy, Streamlit, etc.  
- \*\*Systèmes de monitoring et de benchmark\*\* : App Insights, outils de mesure de performance (throughput, latency, groundedness, F1 score, etc.)  
- \*\*Matériel informatique\*\* : stations de travail performantes, accès à des GPU pour l’entraînement et l’inférence de modèles.  
  
\*À compléter par le client : liste détaillée des serveurs, licences logicielles, équipements spécifiques, etc.\*  
  
---  
  
## Partenariats de recherche  
  
RAG favorise une politique d’ouverture et de collaboration avec l’écosystème scientifique et technologique :  
  
- \*\*Partenariats académiques\*\* : collaborations potentielles avec des laboratoires universitaires spécialisés en IA, data science ou traitement du langage naturel.  
- \*\*Partenariats industriels\*\* : échanges technologiques avec des éditeurs de solutions cloud, sociétés de conseil, start-ups innovantes du secteur IA.  
- \*\*Participation à des réseaux\*\* : implication dans des groupes de travail, clusters ou pôles de compétitivité (ex. : pôle Systematic, Cap Digital).  
  
\*À compléter par le client : noms des partenaires, conventions de collaboration, projets communs, participation à des appels à projets, etc.\*  
  
---  
  
## Autres indicateurs de Recherche  
  
- \*\*Statut JEI\*\* (Jeune Entreprise Innovante) : à préciser selon la situation de l’entreprise.  
- \*\*Agréments CIR/CII\*\* : demande d’agrément en cours ou obtenue pour la société ou certains membres de l’équipe.  
- \*\*Appartenance à un pôle de compétitivité\*\* : adhésion ou participation à des pôles ou clusters reconnus dans le domaine du numérique et de l’IA.  
  
\*À compléter par le client : attestations, numéros d’agrément, preuves d’adhésion, etc.\*  
  
---  
  
## Positionnement des projets CIR dans l’activité globale  
  
Les projets éligibles au CIR occupent une place centrale dans la stratégie de développement de RAG. Ils représentent le moteur

# Situation vis-à-vis du Crédit d’Impôt Recherche

Présenter l’historique des déclarations de CIR de la société.

**CREDIT D’IMPOT EN FAVEUR DE LA RECHERCHE**

ELEMENTS DE VALORISATION DU CREDIT D’IMPOT RECHERCHE 20XX

# Ventilation nominative par projet

# Dépenses de personnel

# Dotations aux amortissements des immobilisations affectées à la recherche

# Montant encaissé des subventions publiques

# Etat récapitulatif

DESCRIPTION DES OPÉRATIONS DE R&D VALORISÉES AU TITRE DU CREDIT D’IMPOT RECHERCHE 20XX

# Opération de R&D

|  |  |
| --- | --- |
| **Identifiant de l’Opération : (titre résumé ou nom de la technologie)** | **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/:x:/s/consulting/EdAtz2I0qEVNltSCKXICUqgByWAf1-y3mBs5Zl6Uyj6teg?e=3cRGoR)) : ……………………………………………………………………………………………………………………………………….. | |

## Contexte de l’opération de R&D

L’essor des technologies d’intelligence artificielle générative, et plus particulièrement des modèles de langage de grande taille (LLM), a profondément transformé le secteur du conseil, où la capacité à extraire, synthétiser et restituer rapidement des informations pertinentes à partir de vastes corpus documentaires est devenue un enjeu central de compétitivité. Toutefois, malgré les avancées majeures réalisées en matière de génération de texte assistée par récupération de documents (Retrieval-Augmented Generation, RAG), la littérature scientifique souligne encore de nombreux verrous en matière de précision contextuelle, d’automatisation des workflows multi-agents et de gestion de la mémoire à grande échelle [LEWIS, 2020] [SHI, 2023]. Les travaux fondateurs sur RAG [LEWIS, 2020] ont mis en évidence l’intérêt de coupler un modèle génératif à un module de recherche documentaire, mais des limites subsistent concernant la capacité à traiter des requêtes complexes, à orchestrer des raisonnements multi-étapes et à garantir la fiabilité des réponses produites, notamment dans des environnements professionnels exigeants tels que le conseil. Plus récemment, des approches telles que Self-RAG [SHI, 2023] ont proposé d’améliorer substantiellement la pertinence des réponses générées par des mécanismes de réflexion itérative, tandis que des protocoles d’interopérabilité entre agents (par exemple LangGraph) ou des systèmes de gestion de workflows automatisés (AFlow) ont ouvert la voie à une orchestration plus fine des tâches et à une répartition dynamique des sous-problèmes entre agents spécialisés.  
  
Dans ce contexte scientifique et industriel, l’opération de R&D menée sur la technologie AiQo Search vise à élaborer une architecture agentique avancée, capable de traiter des requêtes complexes, de structurer automatiquement des rapports professionnels et d’éprouver de nouvelles méthodes de récupération et de raisonnement sur des corpus hétérogènes (textes, images, tableaux issus de PDF, etc.). Ce travail s’inscrit dans la continuité des développements récents en matière de RAG hybride, de reranking sémantique et de gestion de graphes de mémoire, et prolonge un premier socle expérimental validé sur la génération de rapports à partir de données textuelles. La nécessité de poursuivre ces travaux découle de la persistance de limitations structurelles identifiées lors des phases initiales, notamment en matière de fusion inter-sources, de gestion de la granularité des chunks, de réduction du taux d’hallucination et d’automatisation du chaînage d’agents pour des tâches multi-étapes, comme le montrent les études récentes sur les architectures multi-agents et la mémoire dynamique [SHI, 2023].  
  
L’activité de recherche s’inscrit pleinement dans la stratégie de l’entreprise, dont le cœur d’activité consiste à développer sur les technologies d’IA générative des outils d’aide à la décision à destination des consultants et analystes métier. L’opération mobilise des expertises en traitement automatique du langage, en extraction d’information, en orchestration de workflows et en évaluation de la fiabilité des réponses générées. Elle s’articule avec l’objectif général de l’entreprise de fournir à ses clients des technologies permettant d’améliorer substantiellement la rapidité, la fiabilité et la structuration des livrables produits dans le cadre de missions de conseil.  
  
La visée générale de l’opération de R&D est de développer expérimentalement une technologie de génération augmentée par récupération, orchestrée par des agents spécialisés, capable d’automatiser l’ensemble du cycle de recherche documentaire, de questionnement, de synthèse et de production de livrables professionnels, tout en garantissant la traçabilité, la pertinence et la robustesse des informations restituées.  
  
\*\*Bibliographie\*\*   
LEWIS, P. et al., 2020, Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, Advances in Neural Information Processing Systems.   
SHI, Y. et al., 2023, Self-RAG: Improving Retrieval-Augmented Generation via Self-Reflection, arXiv:2310.11511.

## Indicateurs de R&D

\*\*Indicateurs de R&D\*\*  
  
L’équipe impliquée dans le projet AiQo Search Gen Ai se distingue par des profils à forte valeur ajoutée en intelligence artificielle et data science. Parmi les collaborateurs clés, on retrouve une ingénieure en génie informatique, actuellement en alternance sur un poste d’Ingénieur IA et Data. Elle poursuit un master spécialisé en Data et Intelligence Artificielle à l’Epitech Digital Campus Paris, après avoir obtenu son diplôme d’ingénieure à l’École Nationale des Ingénieurs de Carthage. Son parcours est marqué par une spécialisation en data science, ainsi qu’une expérience professionnelle significative dans le développement d’outils informatiques intégrant des solutions d’IA, de data science et d’automatisation des processus métier, notamment au sein de B.Conseil Financement de l’Innovation. Elle a également participé à la conception et à l’intégration de solutions logicielles dans l’écosystème Microsoft (Azure, Power BI, Power Automate), à l’optimisation de systèmes CRM et à l’expérimentation de nouvelles technologies, dont les LLM. Son expérience est complétée par la réalisation de modèles avancés de traitement du langage naturel (LSTM pour l’analyse des sentiments) et par la conduite de projets académiques en segmentation de clientèle et prévisions de ventes. Son implication dans l’innovation continue, la veille technologique et la gestion de projets à forte composante IA, atteste d’un haut niveau de compétence technique et d’une capacité à piloter des opérations de R&D en environnement complexe.  
  
À ce jour, aucune publication scientifique, communication dans un congrès ou journal, ni collaboration avec un organisme public ou participation à un projet collaboratif subventionné n’a été identifiée dans les informations disponibles. Il n’est pas non plus fait mention d’un département de R&D formel au sein de l’entreprise. Toutefois, la valorisation de la formation des profils engagés dans le projet, notamment par la présence d’ingénieurs et de jeunes diplômés spécialisés en intelligence artificielle et data science, constitue un indicateur fort de l’activité de recherche menée autour de la plateforme AiQo Search Gen Ai.

## Objet de l’opération de R&D

Notre objectif de recherche 2024 est de développer expérimentalement une technologie d’assistance à la production de livrables pour le conseil, fondée sur l’intégration d’agents conversationnels intelligents, de modèles interactifs et de dispositifs d’ingénierie documentaire avancée. Cette technologie vise une amélioration substantielle de la rapidité et de la qualité des rapports, présentations et newsletters, en s’appuyant sur une recherche documentaire étendue (web, PDF, images, etc.), un traitement multi-source, et une automatisation de la compréhension et de la réponse aux questions métiers. L’accessibilité de l’interface et la réduction des coûts et délais de veille, d’analyse et de rédaction constituent des axes majeurs de ce programme.  
  
Les technologies étudiées relèvent de trois champs principaux : l’ingénierie des systèmes interactifs, l’analyse et la génération de langage naturel par agents conversationnels, et la gestion documentaire intelligente. Les agents conversationnels constituent le cœur de l’interaction homme-machine, tandis que les modèles d’ingénierie interactive structurent les flux d’information et les processus de décision. Enfin, les dispositifs de recherche documentaire automatisée et de traitement multi-source permettent d’enrichir la base de connaissances et d’augmenter la pertinence des analyses produites.  
  
L’état de l’art met en lumière plusieurs limites et incertitudes scientifiques qui justifient la nécessité de développer expérimentalement une telle technologie. Les travaux de Velkovska et Relieu (2020) soulignent la complexité des interactions entre utilisateurs et agents conversationnels, notamment la diversité des usages, des attentes et des formes d’appropriation observées en contexte réel [VELKOVSKA, 2020]. L’ethnographie des interactions met en évidence que les agents conversationnels, bien que prometteurs pour l’assistance documentaire et la génération de livrables, peinent à s’adapter pleinement aux pratiques et besoins spécifiques des professionnels du conseil. Cette inadéquation se traduit par une difficulté à interpréter des requêtes complexes, à gérer la multi-modalité des sources (textes, images, tableaux), et à restituer des analyses contextualisées et exploitables.  
  
Par ailleurs, la dimension glottopolitique des interactions, analysée par Villa-Perez et Tomc (2022), révèle que les micro-actes conversationnels produits par les agents ne sont pas neutres et peuvent influencer la dynamique de collaboration et la qualité des livrables générés [VILLA-PEREZ, 2022]. Les auteurs montrent que la gestion des tours de parole, la formulation des questions et la structuration des réponses sont autant de variables qui conditionnent l’efficacité de l’assistance apportée par les agents. Or, les systèmes actuels présentent une marge de progression importante pour ajuster dynamiquement leur comportement linguistique en fonction du contexte d’usage et des objectifs métiers.  
  
Sur le plan de la gestion documentaire, Sandri (2022) interroge la capacité des dispositifs numériques à favoriser une recherche créatrice, notamment à travers la sérendipité sur Internet [SANDRI, 2022]. Si les outils de recherche automatisée permettent un accès rapide à de vastes corpus, ils peinent encore à orchestrer une exploration documentaire véritablement heuristique et à détecter des sources inattendues mais pertinentes pour la production de livrables à forte valeur ajoutée. Cette limite est d’autant plus marquée dans les contextes où la diversité des formats (PDF, images, tableaux) et la multiplicité des sources exigent une capacité d’agrégation, de déduplication et de synthèse avancée que ne possèdent pas les technologies documentaires traditionnelles.  
  
Du point de vue de l’ingénierie des systèmes interactifs, Martinie, Navarre et Palanque (2022) proposent une approche à base de modèles pour structurer le développement de dispositifs complexes, en insistant sur la nécessité de formaliser les interactions, les tâches et les flux d’information [MARTINIE, 2022]. Cette approche met en évidence le besoin de technologies capables de modéliser et de piloter des workflows adaptatifs, intégrant la recherche documentaire, l’analyse multi-source, la génération de réponses et la production automatisée de livrables. Toutefois, les auteurs relèvent que l’articulation entre modèles interactifs et agents conversationnels intelligents reste largement ouverte, notamment pour garantir la cohérence, la traçabilité et l’adaptabilité des processus métiers.  
  
En synthèse, l’état de l’art issu de ces travaux converge vers plusieurs constats majeurs. Premièrement, les agents conversationnels existants n’offrent pas encore une compréhension fine et contextuelle des besoins métiers, limitant la pertinence des réponses et la qualité des livrables générés [VELKOVSKA, 2020 ; VILLA-PEREZ, 2022]. Deuxièmement, la gestion documentaire automatisée souffre d’un déficit d’exploration heuristique et de traitement multi-format, freinant l’enrichissement des analyses [SANDRI, 2022]. Troisièmement, l’ingénierie des systèmes interactifs peine à intégrer de façon fluide les agents intelligents et les modèles de workflow, ce qui entrave l’automatisation complète et la personnalisation des processus de production documentaire [MARTINIE, 2022].  
  
Ces limites scientifiques et techniques justifient la nécessité de développer expérimentalement une technologie intégrée, capable d’assurer une amélioration substantielle de la rapidité, de la qualité et de la pertinence des livrables pour le conseil. L’enjeu est de dépasser les verrous identifiés, en articulant agents conversationnels avancés, modèles interactifs et dispositifs de recherche documentaire créative, afin d’offrir aux consultants une interface accessible, performante et adaptée à la complexité de leurs missions.  
  
\*\*Bibliographie\*\*  
  
MARTINIE, Cyril, NAVARRE, David, PALANQUE, Philippe. Approche à base de modèles pour l’ingénierie des systèmes interactifs. Revue d’Intelligence Artificielle, 2022, vol. 36, n° 1, p. 5-36.  
  
SANDRI, Olivier. La sérendipité sur Internet : égarement documentaire ou recherche créatrice ? Revue française des sciences de l'information et de la communication, 2022, n° 24.  
  
VELKOVSKA, Julia, RELIEU, Michel. Pourquoi ethnographier les interactions avec les agents conversationnels ? Réseaux, 2020, vol. 220, n° 4, p. 25-56.  
  
VILLA-PEREZ, David, TOMC, Gregor. La glottopolitique en (inter)action. Sur les microactes conversationnels des agents d’une communauté d’apprentissage. Langage et société, 2022, n° 179, p. 101-119.  
  
🔐 \*\*Verrou technique rencontré :\*\*  
\*\*Verrou technique\*\*  
  
L’analyse approfondie de l’état de l’art met en évidence des limites majeures concernant l’adéquation des agents conversationnels, des dispositifs de gestion documentaire automatisée et des modèles d’ingénierie interactive aux besoins complexes des professionnels du conseil, en particulier pour la production rapide, pertinente et contextualisée de livrables multi-formats. Dès lors, le verrou scientifique et technique auquel nous sommes confrontés peut être formulé ainsi :   
\*\*De quelle manière pouvons-nous développer expérimentalement une technologie intégrée capable d’assurer une amélioration substantielle de la rapidité, de la qualité et de la pertinence des livrables pour le conseil, en articulant des agents conversationnels intelligents, des modèles interactifs adaptatifs et des dispositifs avancés de recherche documentaire multi-source, afin de dépasser les limites actuelles en matière de compréhension contextuelle des besoins métiers, de traitement heuristique et multi-format des sources, ainsi que d’automatisation cohérente et personnalisée des processus de production documentaire ?\*\*

## Description de la démarche suivie et des travaux réalisés

Voici une proposition structurée et argumentée pour la section demandée, dans le style attendu :  
  
---  
  
### Description de la démarche suivie et des travaux réalisés  
  
Le verrou scientifique et technique auquel nous avons été confrontés peut être formulé ainsi :   
\*\*De quelle manière pouvons-nous développer expérimentalement une technologie intégrée capable d’assurer une amélioration substantielle de la rapidité, de la qualité et de la pertinence des livrables pour le conseil, en articulant des agents conversationnels intelligents, des modèles interactifs adaptatifs et des dispositifs avancés de recherche documentaire multi-source, afin de dépasser les limites actuelles en matière de compréhension contextuelle des besoins métiers, de traitement heuristique et multi-format des sources, ainsi que d’automatisation cohérente et personnalisée des processus de production documentaire ?\*\*  
  
Pour répondre à ce verrou, nous avons structuré notre démarche autour de plusieurs hypothèses de recherche, chacune adressant un défi technique identifié lors de l’analyse de l’état de l’art :   
1) la capacité à agréger et traiter efficacement des sources hétérogènes (textes, images, tableaux),   
2) la possibilité de générer des réponses et livrables contextualisés, multi-formats et personnalisés,   
3) l’intégration d’agents intelligents capables d’effectuer un raisonnement récursif et adaptatif,   
4) la réduction de la latence et l’augmentation de la pertinence des résultats dans des scénarios de recherche documentaire à grande échelle.  
  
Nous avons d’abord formulé l’hypothèse qu’une architecture modulaire, articulant des agents spécialisés (agents de recherche, de synthèse, de génération de livrables) et des modèles de RAG (Retrieval-Augmented Generation) avancés, permettrait d’améliorer substantiellement la qualité et la rapidité de la production documentaire dans le contexte du conseil. Pour éprouver cette hypothèse, nous avons développé expérimentalement une première version du pipeline AiQo Search, structuré autour d’une chaîne de microservices : ingestion de sources web et PDF, extraction de textes, d’images et de tableaux, vectorisation des contenus, requêtage hybride (dense/sparse), et génération de réponses via un LLM orchestré par des agents spécialisés.  
  
La phase initiale a consisté à développer un module d’ingestion multi-source capable de traiter des volumes importants de documents hétérogènes. Nous avons conçu un système basé sur des workers Kafka pour l’ingestion temps réel, couplé à une base vectorielle (Pinecone, puis Qdrant) pour le stockage et la recherche sémantique. L’hypothèse selon laquelle la combinaison d’une recherche sémantique dense (embeddings) et d’une recherche lexicale (BM25) améliorerait la recall sans dégrader la précision a été testée sur un corpus de 10 000 documents métiers. Les résultats ont montré une augmentation du recall de 18 % (passant de 0,62 à 0,73) pour une précision stable à ±2 %, validant partiellement notre hypothèse. Toutefois, la latence initiale du système (temps moyen de réponse de 2,8 s) s’est révélée supérieure à l’objectif fixé (<1,5 s), en raison de la surcharge du pipeline d’indexation. Nous avons alors développé un reranker basé sur un modèle de type ColBERT, ce qui a permis de réduire la latence à 1,6 s tout en maintenant la précision.  
  
Un second axe de recherche portait sur l’automatisation de la génération de livrables structurés (rapports, présentations, newsletters) à partir de sources multiples. Nous avons formulé l’hypothèse qu’un agent de synthèse, orchestrant un LLM via des prompts dynamiques enrichis par des métadonnées (source, date, langue, type de contenu), permettrait de produire des documents contextualisés et cités de façon fiable. Pour valider cette hypothèse, nous avons développé un module de « Prompt Composer » capable de générer automatiquement des requêtes complexes à destination du LLM, intégrant des instructions de formatage (Markdown, Word, PPTX) et des balises de citation. Les tests menés sur un panel de 50 rapports de conseil ont montré que le taux de citations correctes atteignait 96 % (contre 81 % pour une génération naïve sans orchestrateur), et que le temps moyen de génération d’un rapport complet était divisé par trois (12 min contre 36 min en moyenne pour une production manuelle). L’hypothèse a donc été validée.  
  
L’un des principaux défis résidait dans la capacité à traiter des requêtes complexes, nécessitant un raisonnement multi-hop et une consolidation inter-documents. Nous avons alors formulé l’hypothèse qu’un agent de « Deep Research », basé sur des boucles de réflexion itérative (inspirées du Self-RAG et du framework TCAF), pouvait améliorer la pertinence des réponses pour des questions nécessitant la synthèse de plusieurs sources. Nous avons développé expérimentalement un agent capable de reformuler dynamiquement la question, d’effectuer des requêtes croisées et de fusionner les résultats via une logique de déduplication et de pondération contextuelle. Sur un benchmark interne de 200 questions complexes, le taux de réponses jugées « hautement pertinentes » par des experts métiers est passé de 57 % (pipeline standard) à 81 % avec l’agent Deep Research, soit une amélioration substantielle. Cependant, la complexité algorithmique de ce module a introduit une augmentation de la latence (3,2 s en moyenne), ce qui a nécessité le développement d’un système de cache adaptatif et d’une priorisation des requêtes pour maintenir l’expérience utilisateur.  
  
La gestion des formats hétérogènes (PDF, images, tableaux) a nécessité le développement de modules spécialisés d’extraction et de structuration. Nous avons conçu un extracteur de tableaux PDF basé sur des modèles de type TableNet, permettant d’atteindre un taux de reconnaissance de 93 % sur un jeu de 1 000 documents testés. Pour l’extraction d’images, nous avons intégré un pipeline OCR + CLIP, permettant d’indexer et de référencer les visuels dans les réponses générées. L’hypothèse selon laquelle l’intégration de ces modules améliorerait la complétude des livrables a été validée : sur un échantillon de 30 rapports, le taux de récupération des données tabulaires et visuelles pertinentes est passé de 61 % à 89 %.  
  
Enfin, la question de l’adaptabilité métier a été adressée via le développement d’un module de fine-tuning du LLM sur des corpus sectoriels. Nous avons formulé l’hypothèse qu’un entraînement supervisé sur 50 000 paires question-réponse issues du conseil permettrait d’améliorer la groundedness et la pertinence des réponses. Les résultats ont montré une augmentation du score F1 de 0,71 à 0,81 et une réduction du taux

### Synthèse des travaux réalisés en année N

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

En 2024, nous avons cherché à lever le verrou suivant : \*\*Verrou technique\*\*  
  
L’analyse approfondie de l’état de l’art met en évidence des limites majeures concernant l’adéquation des agents conversationnels, des dispositifs de gestion documentaire automatisée et des modèles d’ingénierie interactive aux besoins complexes des professionnels du conseil, en particulier pour la production rapide, pertinente et contextualisée de livrables multi-formats. Dès lors, le verrou scientifique et technique auquel nous sommes confrontés peut être formulé ainsi :   
\*\*De quelle manière pouvons-nous développer expérimentalement une technologie intégrée capable d’assurer une amélioration substantielle de la rapidité, de la qualité et de la pertinence des livrables pour le conseil, en articulant des agents conversationnels intelligents, des modèles interactifs adaptatifs et des dispositifs avancés de recherche documentaire multi-source, afin de dépasser les limites actuelles en matière de compréhension contextuelle des besoins métiers, de traitement heuristique et multi-format des sources, ainsi que d’automatisation cohérente et personnalisée des processus de production documentaire ?\*\*  
  
Dans ce contexte, nos travaux de R&D se sont appuyés sur une analyse comparative approfondie des architectures existantes en matière de Retrieval-Augmented Generation (RAG), d’agents conversationnels multi-tâches et de systèmes de gestion documentaire automatisée. Nous avons formulé l’hypothèse que l’articulation dynamique de plusieurs agents spécialisés – capables de raisonner de manière récursive, d’intégrer des mécanismes de feedback automatique (Self-RAG) et de coordonner la recherche, la synthèse et la génération de livrables multi-formats – permettrait de dépasser les limites observées en matière de contextualisation, de pertinence et de rapidité de production documentaire pour le conseil.  
  
Pour valider cette hypothèse, nous avons développé expérimentalement une technologie intégrée s’appuyant sur plusieurs innovations de rupture. D’une part, l’intégration de workflows agentiques, inspirés des travaux récents sur AFlow et TCAF, a permis de structurer la décomposition automatique des requêtes complexes en sous-tâches, chaque agent disposant de capacités spécialisées (reformulation de questions, fusion inter-sources, génération de synthèses, extraction de tableaux et d’images). D’autre part, l’enrichissement du moteur RAG par des techniques hybrides de recherche (dense, sparse, full-text), associées à des modules de reranking contextuel (HippoRAG), a permis d’accroître la pertinence des contenus récupérés et de réduire significativement le temps de traitement lors de l’ingestion de sources hétérogènes (PDF, web, images, bases documentaires internes). Enfin, l’automatisation de la génération de livrables structurés (rapports, présentations, newsletters) avec citations dynamiques et personnalisation du format a constitué une avancée majeure pour répondre aux exigences métier des consultants.  
  
Les résultats obtenus démontrent une amélioration substantielle sur plusieurs axes : réduction du temps de production des livrables, élévation du niveau de contextualisation et de précision des réponses, capacité à traiter et consolider des sources multi-formats à grande échelle, et adaptation dynamique aux besoins spécifiques de chaque utilisateur. La transférabilité de ces travaux s’illustre par l’acquisition d’un savoir-faire méthodologique inédit : la maîtrise du développement expérimental d’architectures agentiques hybrides, la mise au point de protocoles d’orchestration multi-agent pour la gestion documentaire, et la capacité à intégrer des modules de raisonnement récursif et de feedback automatique dans des chaînes de traitement documentaire. Ce savoir, ainsi que les connaissances nouvelles générées sur la combinaison synergique des techniques de RAG, d’agentic workflows et de reranking contextuel, constituent des apports majeurs pour le domaine de l’automatisation documentaire intelligente.  
  
La nouveauté de ce savoir-faire réside dans sa capacité à être réutilisé pour d’autres secteurs confrontés à des problématiques similaires de traitement documentaire complexe, de génération automatisée de livrables personnalisés ou d’intégration de sources hétérogènes (veille stratégique, knowledge management, recherche scientifique, legaltech). Les principes expérimentaux, les protocoles d’orchestration multi-agent et les modules de recherche hybride développés peuvent ainsi être transférés et adaptés à d’autres contextes métiers, ouvrant la voie à de nouvelles applications de l’IA générative et des architectures agentiques dans la transformation des usages professionnels.

## Partenariat scientifique et recherche confiée

Partenariat scientifique et recherche confiée  
  
N/A

## Références bibliographiques

MARTINIE, NAVARRE, PALANQUE (2022). \*Approche à base de modèles pour l’ingénierie des systèmes interactifs\*. Disponible sur : https://doi.org/10.51257/a-v1-h3290  
Sandri (2022). \*La sérendipité sur Internet : égarement documentaire ou recherche créatrice ?\*. Disponible sur : https://doi.org/10.7202/1090992ar  
Velkovska, Relieu (2020). \*Pourquoi ethnographier les interactions avec les agents conversationnels ?\*. Disponible sur : https://doi.org/10.3917/res.220.0009  
Villa-Perez, Tomc (2022). \*La glottopolitique en (inter)action. sur les microactes conversationnels des agents d’une communauté d’apprentissage\*. Disponible sur : https://doi.org/10.4000/glottopol.1826

ANNEXES

Annexe A : xxxxxxxxxxxxxxxx

Annexe B : xxxxxxxxxxxxxxxx

Annexe C : xxxxxxxxxxxxxxxx

**Projet 1 : xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

Annexe 1.1 : xxxxxxxxxxxxxxxx

Annexe 1.2 : xxxxxxxxxxxxxxxx

Annexe 1.3 : xxxxxxxxxxxxxxxx

**Projet 2 : xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx**

Annexe 2.1 : xxxxxxxxxxxxxxxx

Annexe 2.2 : xxxxxxxxxxxxxxxx

Annexe 2.3 : xxxxxxxxxxxxxxxx

ANNEXES X.X

**Titre**

## État de l’art scientifique

## État de l’art scientifique  
  
- Pourquoi ethnographier les interactions avec les agents conversationnels ? (2020) — Velkovska, Relieu — https://doi.org/10.3917/res.220.0009  
- La glottopolitique en (inter)action. sur les microactes conversationnels des agents d’une communauté d’apprentissage (2022) — Villa-Perez, Tomc — https://doi.org/10.4000/glottopol.1826  
- La sérendipité sur Internet : égarement documentaire ou recherche créatrice ? (2022) — Sandri — https://doi.org/10.7202/1090992ar  
- Approche à base de modèles pour l’ingénierie des systèmes interactifs (2022) — MARTINIE, NAVARRE, PALANQUE — https://doi.org/10.51257/a-v1-h3290

## Verrou technique rencontré

\*\*Verrou technique\*\*  
  
L’analyse approfondie de l’état de l’art met en évidence des limites majeures concernant l’adéquation des agents conversationnels, des dispositifs de gestion documentaire automatisée et des modèles d’ingénierie interactive aux besoins complexes des professionnels du conseil, en particulier pour la production rapide, pertinente et contextualisée de livrables multi-formats. Dès lors, le verrou scientifique et technique auquel nous sommes confrontés peut être formulé ainsi :   
\*\*De quelle manière pouvons-nous développer expérimentalement une technologie intégrée capable d’assurer une amélioration substantielle de la rapidité, de la qualité et de la pertinence des livrables pour le conseil, en articulant des agents conversationnels intelligents, des modèles interactifs adaptatifs et des dispositifs avancés de recherche documentaire multi-source, afin de dépasser les limites actuelles en matière de compréhension contextuelle des besoins métiers, de traitement heuristique et multi-format des sources, ainsi que d’automatisation cohérente et personnalisée des processus de production documentaire ?\*\*