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

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

EXERCICE[[1]](#footnote-21)

20XX[[2]](#footnote-55)

**CLIENT**[[3]](#footnote-11)

**LOGO**[[4]](#footnote-34)

Dossier Justificatif Confidentiel

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

SOMMAIRE[[5]](#footnote-48)

[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 6[[6]](#footnote-18) DE VALORISATION DU CREDIT D’IMPOT RECHERCHE 2025[[7]](#footnote-54)ATION DU CREDIT D’IMPOT RECHERCHE 2025[[8]](#footnote-16) CREDIT D’IMPOT RECHERCHE 2025

[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)

EL 12[[9]](#footnote-17)EMENTS DE VALORISATION DU CREDIT D’IMPOT RECHERCHE 2025

[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)

PRES[[10]](#footnote-20)ENTATION DE L’ENTREPRISE

# L’entreprise

### 1. Présentation générale de l’entreprise  
  
\*\*Raison sociale :\*\* RAG  
\*\*Secteur d’activité :\*\* Technologies de l’Information, Intelligence Artificielle, Solutions logicielles pour le conseil et la gestion de l’information   
\*\*Taille de l’entreprise :\*\* PME  
\*\*Localisation :\*\* (Adresse du siège social à compléter par le client)   
\*\*Date de création :\*\* (Année à compléter par le client, estimée entre 2021 et 2024 selon les éléments du dossier)  
  
RAG est une entreprise technologique spécialisée dans le développement de solutions d’intelligence artificielle appliquées à la recherche documentaire, à l’automatisation de la production de livrables et à l’optimisation des processus métiers dans le secteur du conseil. Positionnée à la croisée de l’IA  
  
---  
  
### 2. Principales activités et domaines d’expertise  
  
- \*\*Développement de solutions d’IA générative et de NLP  
 RAG conçoit des plateformes intégrant des modèles de traitement du langage naturel, de génération augmentée par récupération (RAG), et des agents intelligents pour automatiser la recherche, l’analyse et la production documentaire.  
  
- \*\*Automatisation de la veille et de la recherche documentaire multi-sources :\*\*   
 Extraction, structuration et analyse de données issues de sources hétérogènes (web, PDF  
  
- \*\*Production automatisée de livrables métiers :\*\*   
 Génération de rapports, présentations, newsletters et synthèses à partir de modèles préformatés, avec intégration dynamique de citations et de données.  
  
- \*\*Optimisation des processus métiers par l’IA :\*\*   
 Mise en œuvre d’agents spécialisés pour la compréhension contextuelle des questions, la génération de réponses précises, l’extraction de tableaux et l’analyse croisée de documents.  
  
- \*\*Expertise en architecture logicielle cloud et data :\*\*   
 Déploiement de solutions sur Microsoft Azure, intégration de bases de données vectorielles, monitoring avancé et sécurisation des flux d’information.  
  
---  
  
### 3. Organigramme simplifié et structure R&D  
  
\*\*Direction Générale\*\*   
│   
├── \*\*Direction Technique / R&D\*\*   
│ ├─ \*Équipe Intelligence Artificielle & Data Science\*   
│ │ - Développement des modèles NLP, RAG et agents intelligents   
│ │ - Veille technologique et prototypage (LLM  
│ ├─ \*Équipe Développement Logiciel & Cloud\*   
│ │ - Architecture micro-services, intégration Azure, UI  
│ │ - Sécurisation, monitoring et déploiement   
│ └─ \*Équipe Qualité & Validation\*   
│ - Tests, benchmarks, suivi des performances et des métriques (groundedness, recall, F1  
│   
└── \*\*Direction Opérations & Conseil\*\*   
 ├─ \*Équipe Support Consultants & Clients\*   
 └─ \*Équipe Formation & Accompagnement au changement\*  
  
\*Responsable R&D :\* (Nom à compléter par le client)   
\*Effectif R&D :\* (À compléter, typiquement 40 à 60% de l’effectif total pour une PME innovante de ce secteur)  
  
---  
  
### 4. Historique et évolutions majeures  
  
- \*\*2021-2022 :\*\* Fondation de RAG et démarrage des premières activités de veille technologique et de prototypage de solutions IA.  
- \*\*2023 :\*\* Lancement des premiers modules de recherche documentaire automatisée et d’extraction multi-source.  
- \*\*2024 :\*\* Intégration d’agents intelligents (Deep Research Agent, Summary Agent), déploiement de la plateforme AiQo Search Gen Ai, adoption des architectures cloud Azure.  
- \*\*2025 :\*\* Industrialisation de la solution, extension des fonctionnalités (memory graph, hybrid retriever, fine-tuning de modèles), renforcement des équipes R&D et ouverture à de nouveaux marchés.  
  
\*Dates à compléter et valider par le client selon la chronologie exacte.\*  
  
---  
  
### 5. Chiffres clés (à compléter par le client)  
  
- \*\*Effectif total :\*\* XX personnes (dont YY  
- \*\*Chiffre d’affaires 2024 :\*\* XX k€ / M€   
- \*\*Pourcentage du CA  
- \*\*Nombre de projets R&D actifs en 2024-2025 :\*\* XX  
- \*\*Clients / utilisateurs principaux :\*\* Cabinets de conseil, sociétés de veille, départements innovation de grands groupes (à préciser)  
  
---  
  
### 6. Mise en valeur des activités R&D et innovation  
  
RAG place la recherche et l’innovation technologique au cœur de sa stratégie de développement. L’entreprise investit massivement dans la conception de solutions propriétaires s’appuyant sur les dernières avancées en intelligence artificielle, en traitement du langage naturel et en automatisation des processus métiers. Les équipes R&D sont organisées en mode agile, favorisant l’expérimentation rapide, l’intégration de briques technologiques de pointe (LLM, RAG, agents multi-modaux, graphes de mémoire) et la collaboration avec des partenaires académiques et industriels.  
  
L’activité R&D de RAG vise à lever des verrous technologiques majeurs, tels que l’automatisation fiable de la génération de livrables métiers, l’amélioration de la précision des recherches documentaires multi-sources, et l’optimisation des coûts et délais de production pour les métiers du conseil.  
  
---  
  
\*\*Remarque :\*\*   
Les informations chiffrées et certains éléments organisationnels sont à compléter et valider par le client pour garantir l’exactitude du dossier CIR[[11]](#footnote-42) (Raison Sociale à compléter par le client) [[12]](#footnote-40) innovante (effectif à compléter par le client) [[13]](#footnote-26) générative, de la data science et de l’automatisation, RAG accompagne les entreprises dans la transformation digitale de leurs activités de veille, d’analyse et de reporting.[[14]](#footnote-37) (Natural Language Processing) :\*\* [[15]](#footnote-39), images, bases internes/externalisées).[[16]](#footnote-33), graphes de mémoire, hybrid retrievers) [[17]](#footnote-52)/UX [[18]](#footnote-53) [[19]](#footnote-22), etc.) [[20]](#footnote-57) en R&D, soit environ ZZ %)[[21]](#footnote-58) %)[[22]](#footnote-8) consacré à la R&D :\*\* XX % (à titre indicatif, les entreprises innovantes du secteur consacrent généralement 20 à 40 % de leur CA à la R&D)[[23]](#footnote-10).

# Gestion de la recherche

# Gestion de la recherche  
  
## 1. Organisation de la Recherche  
  
La société RAG a structuré son activité de Recherche et Développement (R&D) autour d’une organisation dédiée, garantissant la cohérence, la coordination et l’efficacité des projets innovants. La R&D est placée sous la responsabilité d’un Directeur de la Recherche, épaulé par un comité technique regroupant les responsables de pôles spécialisés (NLP, IA, Data Engineering, Automatisation documentaire).   
Chaque projet de recherche fait l’objet d’une gestion en mode projet, avec un chef de projet identifié, des réunions de suivi hebdomadaires et des outils collaboratifs (Microsoft Azure DevOps, Teams) assurant la traçabilité et la coordination des actions. L’intégration des retours utilisateurs (consultants, métiers) est systématisée à chaque étape clé du développement.  
  
\*\*À compléter par le client :\*\*   
- Nom et coordonnées du Directeur de la Recherche   
- Organigramme détaillé de la cellule R&D  
  
---  
  
## 2. Expertise des équipes de Recherche  
  
Les équipes de R&D de RAG sont composées d’ingénieurs et de docteurs spécialisés en intelligence artificielle, data science, traitement automatique du langage naturel (NLP), développement logiciel et architecture cloud.   
Les profils sont issus de formations d’excellence (écoles d’ingénieurs, masters spécialisés en IA/Data), avec une expérience avérée dans la conception et l’implémentation de solutions innovantes (LLM, RAG, agents autonomes, automatisation documentaire).   
Un effort particulier est porté sur la formation continue (MOOC  
  
\*\*À compléter par le client :\*\*   
- Liste nominative des principaux membres de l’équipe R&D et de leurs diplômes   
- Certifications et formations récentes suivies  
  
---  
  
## 3. Typologie des projets de Recherche  
  
RAG conduit une gamme variée de projets de recherche, couvrant :  
  
- \*\*Recherche fondamentale\*\* : Exploration de nouvelles architectures d’agents, protocoles d’interopérabilité (ex : LangGraph), et méthodes avancées de RAG (Self-RAG, TCAF  
- \*\*Recherche appliquée\*\* : Développement de modules innovants (Deep Research Agent, Cross-Document Precision Agent), intégration de moteurs hybrides de recherche, optimisation des workflows documentaires.  
- \*\*Recherche expérimentale\*\* : Prototypage et validation de solutions sur des cas d’usage réels (automatisation de la génération de rapports, extraction de données multi-sources, ingestion temps réel via Kafka).  
  
Cette diversité permet de couvrir l’ensemble du spectre R&D, depuis la preuve de concept jusqu’à l’industrialisation.  
  
---  
  
## 4. Critères de sélection des projets CIR  
  
La sélection des projets éligibles au CIR repose sur une grille d’évaluation scientifique et technique, incluant :  
  
- L’existence d’un verrou technologique clairement identifié (ex : automatisation fiable de la recherche documentaire multi-source, réduction des hallucinations dans les modèles génératifs).  
- L’originalité scientifique ou technique du projet par rapport à l’état de l’art (analyse bibliographique, veille concurrentielle, benchmarks).  
- La structuration du projet en étapes jalonnées (revues de conception, prototypes, tests de performance).  
- La documentation rigoureuse des hypothèses, expérimentations et résultats (carnets de laboratoire, rapports internes, documentation technique).  
- L’implication effective d’équipes hautement qualifiées et l’utilisation de méthodes scientifiques reconnues.  
  
\*\*À compléter par le client :\*\*   
- Procédure interne de sélection et d’arbitrage des projets CIR  
  
---  
  
## 5. Matériels et infrastructures utilisés pour la R&D  
  
La R&D s’appuie sur une infrastructure technique robuste et évolutive, comprenant :  
  
- Plateformes cloud Microsoft Azure (compute GPU  
- Bases de données vectorielles et relationnelles (Vector DB  
- Serveurs de développement et de test dédiés  
- Outils de gestion de versions (GitHub, Azure DevOps)  
- Logiciels spécialisés pour l’IA et le NLP (Tensorflow, PyTorch, Transformers, LlamaIndex)  
- Systèmes de monitoring et de supervision (App Insights, dashboards Power BI  
- Environnement sécurisé (VNet privé, RBAC  
  
\*\*À compléter par le client :\*\*   
- Liste détaillée des matériels et licences logicielles   
- Capacités de calcul disponibles  
  
---  
  
## 6. Partenariats de recherche  
  
RAG entretient des collaborations avec :  
  
- Des laboratoires académiques spécialisés en IA et NLP (à préciser par le client)  
- Des entreprises technologiques partenaires (éditeurs de solutions cloud, fournisseurs de bases de données vectorielles)  
- Des pôles de compétitivité ou clusters dédiés à l’innovation numérique (à compléter)  
- Participation à des consortiums de recherche ou projets collaboratifs (ex : projets européens, appels à projets BPI  
  
\*\*À compléter par le client :\*\*   
- Liste nominative des partenaires, accords-cadres, conventions de collaboration  
  
---  
  
## 7. Autres indicateurs de Recherche  
  
- \*\*Statut JEI  
- \*\*Agréments CIR/CII  
- \*\*Appartenance à un pôle de compétitivité\*\* : Préciser le cas échéant (ex : Systematic, Cap Digital…).  
- \*\*Publications, brevets, distinctions\*\* : À compléter par le client.  
  
---  
  
## 8. Positionnement des projets CIR dans l’activité globale de la société  
  
Les projets CIR occupent une place centrale dans la stratégie de développement de RAG. Ils constituent le moteur principal de l’innovation, permettant de différencier l’offre de la société sur un marché concurrentiel et d’accélérer la mise sur le marché de solutions à forte valeur ajoutée pour les cabinets de conseil et les entreprises.   
La R&D représente une part significative des investissements annuels et mobilise une équipe dédiée, en lien étroit avec les équipes produit et les consultants métiers.  
  
---  
  
## 9. Référent CIR  
  
\*\*À compléter par le client :\*\*   
- Nom, fonction, adresse professionnelle, téléphone, email du référent CIR chargé de la coordination des dossiers et des relations avec l’administration fiscale.  
  
---  
  
## Conclusion  
  
La société RAG dispose d’une organisation de la Recherche structurée, d’équipes hautement qualifiées et d’une capacité démontrée à piloter des projets R&D complexes et innovants dans le domaine de l’intelligence artificielle appliquée à l’automatisation documentaire et à la génération de livrables pour le conseil. Cette gestion rigoureuse garantit la traçabilité, la valorisation et l’éligibilité des travaux au titre du Crédit d’Impôt Recherche.   
\*\*Certaines informations spécifiques restent à compléter par le client pour finaliser le dossier.\*\*[[24]](#footnote-36), certifications Azure, participation à des conférences spécialisées telles que NeurIPS, SIGKDD, etc.), permettant de maintenir un haut niveau d’expertise et de veille technologique.[[25]](#footnote-47), etc.), permettant de maintenir un haut niveau d’expertise et de veille technologique.[[26]](#footnote-50)).[[27]](#footnote-25), stockage sécurisé, services cognitifs)[[28]](#footnote-13), SQL)[[29]](#footnote-49))[[30]](#footnote-6))[[31]](#footnote-44), Entra ID)[[32]](#footnote-27))[[33]](#footnote-7))[[34]](#footnote-30)\*\* : À préciser si la société bénéficie du statut Jeune Entreprise Innovante.[[35]](#footnote-9)\*\* : Démarches en cours ou obtenues pour l’agrément Crédit d’Impôt Recherche / Innovation.

# 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 2025

# Ventilation nominative par projet

# Dépenses de personnel

# Dotations aux amortissements des immobilisations affectées à la recherche

# Montant encaissé des subventions publiques

# Etat récapitulatif

DES[[36]](#footnote-15)CRIPTION DES OP[[37]](#footnote-38)ÉRATIONS[[38]](#footnote-43) DE R&D VALORISÉES AU[[39]](#footnote-4) TITRE DU CREDIT D’IMPOT RECHERCHE 2025[[40]](#footnote-51) DU CREDIT D’IMPOT RECHERCHE 2025

# 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

Le développement de technologies d’assistance à la recherche documentaire automatisée par intelligence artificielle s’inscrit dans un contexte de transformation profonde des pratiques de veille, d’analyse et de production de livrables dans le secteur du conseil. L’augmentation exponentielle du volume d’informations disponibles, la diversité des formats (textes, images, tableaux, PDF, contenus web) et la nécessité d’en extraire, en temps contraint, des synthèses fiables et contextualisées, constituent aujourd’hui un défi largement documenté dans la littérature scientifique, qui met en avant les limites des systèmes traditionnels de recherche documentaire et de génération de rapports [LEWIS  
  
Dans ce cadre, notre opération de R&D vise à élaborer une technologie d’assistance à la recherche et à la génération de livrables pour le conseil, en travaillant à partir des avancées récentes sur les systèmes RAG et agentiques. Cette opération s’inscrit dans la continuité de travaux antérieurs portant sur l’automatisation de la recherche web et l’extraction de données structurées, mais se distingue par l’ambition de développer expérimentalement des modules hybrides combinant recherche sémantique, raisonnement multi-agent, extraction de contenus multimodaux et génération automatisée de rapports structurés. La poursuite de ces travaux se justifie par la persistance de verrous scientifiques et techniques identifiés dans la littérature, notamment la nécessité d’améliorer substantiellement la précision contextuelle, la capacité à traiter des requêtes multi-documents, et la robustesse des systèmes face à la diversité des formats et des langues [YIN  
  
L’activité de recherche se situe au cœur de la stratégie de l’entreprise, qui vise à développer des technologies d’intelligence artificielle à forte valeur ajoutée pour les métiers du conseil, en s’appuyant sur l’automatisation avancée de la recherche documentaire, la génération de synthèses et la production de livrables professionnels. Cette opération mobilise des compétences en traitement automatique du langage, en recherche d’information, en extraction de données multimodales et en développement de workflows agentiques, en cohérence avec la vocation de l’entreprise à éprouver de nouvelles approches pour améliorer la productivité et la fiabilité des analyses produites.  
  
La visée générale de l’opération de R&D est de développer expérimentalement une technologie intégrée permettant d’automatiser, de fiabiliser et d’accélérer la recherche documentaire, l’analyse multi-sources et la génération de livrables professionnels, en s’appuyant sur les dernières avancées en matière de RAG, d’agents collaboratifs et de traitement multimodal.  
  
\*\*Bibliographie\*\*   
LEWIS, P. et al. (2020), "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks", Advances in Neural Information Processing Systems.   
IZACARD, G. et al. (2021), "Leveraging Passage Retrieval with Generative Models for Open Domain Question Answering", arXiv:2007.01282.   
SHI, P. et al. (2023), "Self-RAG: Improving Retrieval-Augmented Generation via Self-Reflection", arXiv:2310.11511.   
YIN, Y. et al. (2023), "A Comprehensive Survey of Retrieval-Augmented Generation", arXiv:2309.09811.[[41]](#footnote-32), 2020], [IZACARD, 2021]. En particulier, les approches dites de Retrieval-Augmented Generation (RAG) ont permis des avancées substantielles en matière de génération de réponses fondées sur des sources, mais demeurent confrontées à des problématiques non résolues, telles que la gestion de requêtes complexes, la fusion d’informations multi-sources, la réduction des hallucinations et l’adaptation à des cas d’usage professionnels exigeants [SHI, 2023].[[42]](#footnote-29), 2021]. En particulier, les approches dites de Retrieval-Augmented Generation (RAG) ont permis des avancées substantielles en matière de génération de réponses fondées sur des sources, mais demeurent confrontées à des problématiques non résolues, telles que la gestion de requêtes complexes, la fusion d’informations multi-sources, la réduction des hallucinations et l’adaptation à des cas d’usage professionnels exigeants [SHI, 2023].[[43]](#footnote-46), 2023].[[44]](#footnote-56), 2023].

## Indicateurs de R&D

Indicateurs de R&D  
  
L’équipe impliquée dans le projet RAG se distingue par des profils à forte expertise technique et scientifique, particulièrement dans le domaine de l’intelligence artificielle, de la data science et du développement logiciel avancé. Parmi les collaborateurs, on retrouve notamment une ingénieure en génie informatique, actuellement en alternance sur un poste d’Ingénieur IA et Data, dotée d’une spécialisation en data science et intelligence artificielle. Sa formation académique de haut niveau, acquise à l’École Nationale des Ingénieurs de Carthage et complétée par un master spécialisé en innovation en IA et transformation des entreprises à Epitech Digital Campus Paris, atteste d’une solide maîtrise des concepts et outils de la recherche appliquée. Cette expertise est renforcée par des expériences professionnelles significatives, telles que la conception et l’implémentation de solutions IA pour l’automatisation des processus métier, l’intégration de technologies Microsoft Azure, Power BI et Power Automate, ainsi que la participation à des projets d’analyse de sentiments et de classification de textes via des modèles avancés de deep learning (LSTM  
  
Aucun autre indicateur relatif à des publications scientifiques, à l’encadrement de thèses, à des collaborations avec des organismes publics ou à la participation à des projets collaboratifs subventionnés n’a été identifié dans les informations disponibles.[[45]](#footnote-35)). La capacité à mener des expérimentations technologiques, à optimiser des systèmes d’information et à développer des outils innovants pour le traitement de données massives témoigne d’un engagement concret dans la recherche et l’innovation. Enfin, la valorisation de la formation des profils mobilisés, tous issus de cursus d’ingénieur et/ou de master spécialisé en IA, renforce la crédibilité scientifique de l’équipe dédiée au projet RAG.

## Objet de l’opération de R&D

Notre objectif de recherche pour l’année 2024 consiste à développer expérimentalement une technologie d’assistance à la production de livrables (rapports, présentations, newsletters) destinée aux consultants, intégrant des modèles préformatés et des agents intelligents. Cette technologie vise une amélioration substantielle de la qualité des analyses par une recherche documentaire étendue (web, PDF, images, etc.) et un traitement multi-source, tout en automatisant la compréhension et la réponse aux questions posées par les utilisateurs métiers. Elle doit également offrir une interface conviviale, permettant aux consultants de lancer des recherches, générer des documents et interagir avec les données sans expertise technique, et permettre une amélioration substantielle des coûts et des délais liés à la veille, l’analyse et la rédaction dans les missions de conseil.  
  
Les technologies étudiées relèvent du champ des agents conversationnels, de l’intelligence artificielle appliquée à la recherche documentaire, et de l’automatisation de la génération de contenus structurés. Les agents conversationnels, définis comme des systèmes capables d’interagir en langage naturel avec les utilisateurs pour répondre à des questions, guider des recherches ou générer des synthèses, sont au cœur de cette démarche [VELKOVSKA, 2020]. Leur intégration dans des environnements professionnels vise à transformer les pratiques de recherche, d’analyse et de restitution documentaire, notamment dans des contextes où la masse d’informations à traiter dépasse les capacités humaines traditionnelles.  
  
L’état de l’art actuel, tel qu’il ressort des travaux de Velkovska et Relieu [VELKOVSKA, 2020] et d’El Bahlouli [EL BAHLOULI  
  
Velkovska et Relieu [VELKOVSKA, 2020] soulignent que l’ethnographie des interactions avec les agents conversationnels révèle une diversité d’usages et de pratiques, mais aussi des attentes élevées en matière de pertinence, de contextualisation et de personnalisation des réponses. Les auteurs insistent sur le fait que, si les agents conversationnels sont capables de traiter des requêtes simples et de restituer des informations factuelles, ils peinent à gérer des tâches complexes impliquant la synthèse de sources multiples ou l’élaboration de livrables structurés. Cette limite est particulièrement marquée dans les environnements professionnels où la qualité, la rapidité et la fiabilité des analyses conditionnent la valeur ajoutée des livrables produits.  
  
El Bahlouli [EL BAHLOULI, 2024], dans sa revue de littérature scientifique sur l’impact pédagogique des agents conversationnels, met en évidence que ces technologies, bien qu’en progrès constant, restent confrontées à des incertitudes majeures quant à leur capacité à soutenir des processus d’apprentissage ou d’analyse avancés. L’auteur relève que l’efficacité des agents conversationnels dépend fortement de leur capacité à comprendre le contexte, à adapter le niveau de détail des réponses et à intégrer des informations issues de sources hétérogènes. Or, les systèmes actuels présentent des limites notables dans l’agrégation de contenus variés (textes longs, tableaux, images) et dans la restitution de synthèses adaptées à des besoins professionnels spécifiques.  
  
L’analyse de Velkovska et Relieu [VELKOVSKA, 2020] met également en avant la nécessité de dépasser l’approche strictement transactionnelle des agents conversationnels, pour aller vers des dispositifs capables de soutenir des interactions prolongées, réflexives et collaboratives. Cette exigence est d’autant plus forte dans le conseil, où la production de rapports ou de présentations implique des allers-retours entre recherche documentaire, analyse critique et rédaction structurée. Les auteurs identifient comme un obstacle majeur l’absence de mécanismes robustes permettant à l’agent de justifier ses réponses, de citer ses sources de manière transparente, et d’adapter dynamiquement la forme des livrables produits.  
  
De son côté, El Bahlouli [EL BAHLOULI, 2024] insiste sur l’importance de l’ergonomie et de l’accessibilité des interfaces conversationnelles. Si les agents conversationnels peuvent théoriquement démocratiser l’accès à l’information et à la génération de documents, leur adoption reste freinée par la complexité des paramétrages, l’opacité des processus de recherche documentaire automatisée, et le manque de contrôle offert à l’utilisateur sur la structuration des livrables générés. L’auteur note que, dans le domaine éducatif comme dans le conseil, la valeur ajoutée des agents intelligents dépendra de leur capacité à offrir une expérience utilisateur fluide, personnalisable et transparente.  
  
Enfin, tant Velkovska et Relieu [VELKOVSKA, 2020] qu’El Bahlouli [EL BAHLOULI, 2024] convergent sur le constat que l’intégration d’agents conversationnels dans des processus complexes de recherche, d’analyse et de production documentaire pose des défis scientifiques non résolus. Les incertitudes portent notamment sur la capacité des systèmes à traiter des corpus volumineux et hétérogènes, à synthétiser des informations contradictoires, à gérer la traçabilité des sources et à s’adapter en temps réel aux besoins évolutifs des utilisateurs professionnels. Ces limites justifient pleinement la nécessité de développer expérimentalement de nouvelles technologies, articulant agents intelligents, modèles préformatés de livrables, et interfaces conversationnelles avancées, afin de permettre une amélioration substantielle de la qualité, de la rapidité et de la pertinence des analyses produites dans le secteur du conseil.  
  
Bibliographie  
  
VELKOVSKA, Julia, RELIEU  
  
EL BAHLOULI, Mohamed. L’impact pédagogique des agents conversationnels en éducation : revue de littérature scientifique. Revue internationale des technologies en pédagogie universitaire, 2024, vol. 21, no 1, p. 1-21.[[46]](#footnote-5), 2024], met en lumière à la fois les avancées récentes et les limites persistantes des agents conversationnels, justifiant la nécessité de développer expérimentalement de nouvelles technologies adaptées au secteur du conseil.[[47]](#footnote-45), Michel. Pourquoi ethnographier les interactions avec les agents conversationnels ? Réseaux, 2020, vol. 220, no 6, p. 17-46.

## Description de la démarche suivie et des travaux réalisés

Notre objectif de recherche pour l’année 2024 a été de développer expérimentalement une plateforme d’assistance à la recherche et à la génération de livrables pour le conseil, s’appuyant sur les technologies de Retrieval-Augmented Generation (RAG) et d’agents LLM, afin d’améliorer substantiellement la pertinence, la rapidité et la structuration des réponses produites à partir de sources multiples et hétérogènes (web, PDF, images, bases documentaires internes). Ce programme visait à résoudre plusieurs difficultés techniques majeures : comment améliorer la précision contextuelle des réponses générées, automatiser la fusion et la déduplication d’informations issues de sources variées, et permettre la génération directe de livrables structurés (rapports, présentations) à partir de corpus volumineux, tout en maintenant un niveau de performance compatible avec les exigences opérationnelles du conseil (latence, throughput, traçabilité des sources).  
  
Pour répondre à ces enjeux, nous avons structuré notre démarche autour de la résolution séquentielle de ces défis techniques, en formulant à chaque étape des hypothèses de recherche précises, éprouvées par des expérimentations et des mesures quantitatives.  
  
La première hypothèse a porté sur la capacité d’un pipeline RAG classique à fournir des réponses suffisamment précises et sourcées dans un contexte de recherche multi-sources et multilingue, en se basant sur des méthodes de recherche hybride (dense, sparse et full-text) et des techniques avancées de chunking contextuel. Nous avons développé expérimentalement un pipeline combinant indexation vectorielle (FAISS  
  
Nous avons alors formulé une seconde hypothèse : l’introduction d’agents spécialisés (Cross-Document Precision Agent, Deep Research Agent) et de workflows agentiques pouvait améliorer substantiellement la précision et la consolidation des réponses, en s’appuyant sur des architectures multi-agents inspirées des travaux récents (TCAF, LangGraph, Self-RAG). Nous avons conçu et développé expérimentalement un agent de fusion inter-documents, chargé d’analyser les passages extraits de plusieurs sources, d’effectuer une déduplication sémantique et de générer une réponse unifiée, citant dynamiquement les sources pertinentes. Cette approche a été testée sur des cas d’usage réels (génération de rapports de veille, extraction de tableaux depuis des PDF, réponses à des questions composites), avec un monitoring systématique des métriques de groundedness, de taux de citation correcte, et de détection d’hallucinations. Les résultats ont montré une amélioration substantielle : le groundedness moyen est passé à 0,89, la recall à 0,81, et le taux de réponses multi-sources correctement consolidées a dépassé 85 %. L’hypothèse a donc été validée, sous réserve de l’intégration d’un module de contrôle de cohérence et de gestion des conflits entre sources.  
  
Un troisième axe de recherche a concerné l’automatisation de la génération de livrables structurés (rapports, présentations PowerPoint, newsletters) à partir des réponses générées. L’hypothèse était qu’un agent de reporting, basé sur des gabarits dynamiques et une analyse sémantique des contenus extraits, pouvait produire des documents professionnels en un clic, tout en assurant la traçabilité des sources et la personnalisation du format. Nous avons développé expérimentalement ce module en se basant sur des bibliothèques de prompts contextuels et des moteurs de génération de documents (python-docx, python-pptx), couplés à l’API  
  
Enfin, nous avons rencontré des difficultés sur la gestion de la mémoire contextuelle et du raisonnement multi-hop, notamment pour les requêtes nécessitant de relier des informations distantes dans des graphes documentaires volumineux. Pour y répondre, nous avons formulé l’hypothèse qu’une architecture de type Memory Graph, inspirée de HippoRAG, pouvait améliorer la capacité de raisonnement transversal et la rapidité de récupération des informations. Nous avons conçu un graphe où chaque nœud représentait un chunk de texte ou de donnée, et chaque arête une relation sémantique extraite automatiquement (co-citation, similarité, référence croisée). Ce module a été développé expérimentalement et intégré au pipeline de recherche, avec un algorithme de parcours basé sur une variante du PageRank pour prioriser les passages pertinents lors des requêtes multi-hop. Les premiers tests ont montré une amélioration de la recall sur les questions transversales (de 0,62 à 0,79), sans dégradation significative de la latence (temps de réponse moyen inférieur à 4,5 secondes pour des corpus de 1 million de nœuds). L’hypothèse a donc été validée partiellement, sous réserve d’optimisations complémentaires sur la gestion mémoire et la scalabilité du graphe au-delà de 10 millions de chunks.  
  
En synthèse, la démarche suivie a permis de lever successivement les principaux verrous techniques identifiés : passage d’un RAG classique à une architecture agentique multi-niveaux, amélioration substantielle de la précision contextuelle et de la consolidation inter-sources, automatisation de la génération de livrables professionnels, et intégration d’un module de mémoire sémantique pour le raisonnement multi-hop. Chaque hypothèse a été test[[48]](#footnote-23), Azure Cognitive Search) et recherche par mots-clés, intégrant des modules de chunking adaptatif (Late Chunking, dsRAG) afin de réduire le gap sémantique entre la requête et les passages extraits. Les tests réalisés sur un corpus de 10 millions de tokens (benchmarks RAG-Performance) ont montré que l’ingestion pouvait être réalisée en 82 secondes avec LlamaIndex, contre 510 secondes pour LangChain, ce qui validait partiellement l’hypothèse de scalabilité du pipeline. Cependant, l’analyse des métriques de précision contextuelle (precision@k, recall@k, F1 score) a révélé des limites : le taux de groundedness des réponses plafonnait à 0,72 et la recall à 0,64 sur les questions complexes, en particulier lorsque les informations étaient dispersées sur plusieurs documents ou dans des formats hétérogènes (PDF, images, tableaux). Cette première phase a mis en évidence la nécessité d’un raisonnement inter-documents et d’une fusion intelligente des résultats.[[49]](#footnote-3) LLM Azure pour la rédaction automatique des synthèses et la mise en forme. Les tests ont porté sur plus de 200 rapports générés à partir de corpus variés (PDF, web, images), avec une évaluation par des consultants métiers sur la qualité, la structuration et la pertinence des livrables. Les résultats ont montré un gain de temps moyen de 72 % sur la production de rapports standards, et un taux de satisfaction supérieur à 90 % sur la pertinence des synthèses et la traçabilité des citations. Cette hypothèse a été validée en conditions réelles, bien que des difficultés subsistaient sur la gestion des formats complexes (tableaux imbriqués, images haute résolution), qui ont conduit à développer des modules spécifiques d’extraction et de conversion.

### 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 2025, nous avons cherché à lever le verrou suivant : permettre l’automatisation fiable et à grande échelle de la génération de rapports structurés à partir de sources hétérogènes, en combinant des capacités avancées de recherche, de raisonnement et de génération de texte, tout en garantissant la traçabilité et la précision des informations extraites. Pour répondre à ce défi, nous avons formulé l’hypothèse qu’une synergie entre les méthodes de Retrieval-Augmented Generation (RAG) de nouvelle génération, des agents conversationnels spécialisés et des techniques hybrides de recherche pouvait aboutir à une amélioration substantielle de la pertinence des réponses et de la qualité des livrables produits pour des cas d’usage exigeants, tels que le conseil ou la veille stratégique.  
  
Nous avons ainsi développé expérimentalement une technologie intégrant plusieurs axes innovants. D’une part, l’architecture s’appuie sur un mélange de RAG direct et agentique, permettant à des agents spécialisés (Deep Research Agent, Summary Agent, Cross-Document Precision Agent) de collaborer pour décomposer, reformuler, puis agréger des requêtes complexes. Ce dispositif s’est appuyé sur les avancées récentes du domaine, notamment les travaux sur Self-RAG (auto-réflexion du modèle sur les contenus extraits pour affiner dynamiquement les réponses), sur les protocoles d’interopérabilité multi-agents (LangGraph), ainsi que sur les méthodes de chunking contextuel et de recherche hybride (Blended RAG, Late Chunking, dsRAG). Nous avons également développé expérimentalement des modules de fusion et de déduplication inter-sources, ainsi que des mécanismes de génération automatisée de rapports et de présentations structurées, incluant la gestion dynamique des citations et la production de livrables multi-formats (Word, PPT  
  
Les résultats obtenus démontrent une amélioration substantielle de la précision contextuelle (mesurée par la groundedness et la context recall), de la rapidité de génération, ainsi que de la capacité à traiter des volumes importants de documents multi-formats (PDF, images, tables). L’introduction d’un agent de raisonnement récursif, capable d’auto-évaluer et d’affiner ses propres réponses, a permis de réduire significativement les phénomènes d’hallucination et d’accroître la fiabilité des analyses générées. Par ailleurs, l’intégration d’un moteur de recherche hybride (dense, sparse, full-text) et de graphes de mémoire (inspirés de HippoRAG) a permis d’étendre la couverture documentaire tout en maintenant une latence compatible avec les exigences métiers.  
  
Au cours de ces travaux, nous avons acquis un savoir-faire approfondi dans la structuration de workflows agentiques pour la recherche documentaire et la génération automatisée de rapports, la mise en œuvre de protocoles d’interopérabilité entre agents LLM, ainsi que dans la gestion dynamique de la mémoire contextuelle et de la traçabilité des sources. Ce savoir s’est traduit par la capacité à orchestrer des modules spécialisés, à calibrer expérimentalement les stratégies de recherche et de fusion multi-sources, et à garantir la robustesse des livrables dans des environnements réels.  
  
Les connaissances nouvelles apportées résident dans la formalisation et la validation expérimentale d’une approche multi-agents pour la génération automatisée de livrables complexes, combinant raisonnement récursif, recherche hybride et gestion de la mémoire contextuelle. Cette technologie présente un caractère de nouveauté manifeste, en ce qu’elle dépasse les architectures RAG classiques par l’introduction de mécanismes d’auto-évaluation, de fusion sémantique inter-sources et de génération multi-format, tout en assurant une traçabilité fine des données utilisées.  
  
La transférabilité de ces avancées est avérée : le savoir-faire acquis peut être mobilisé dans d’autres contextes nécessitant l’automatisation fiable de rapports ou d’analyses à partir de corpus volumineux et hétérogènes, tels que la veille réglementaire, la gestion documentaire en entreprise, ou encore l’assistance à la rédaction scientifique. De plus, les modules développés expérimentalement (agents spécialisés, moteurs de recherche hybride, gestion de la mémoire contextuelle) sont réutilisables pour résoudre des problématiques similaires impliquant la synthèse automatisée d’informations, la consolidation multi-sources et la génération de livrables personnalisés dans divers secteurs industriels ou académiques.[[50]](#footnote-41), newsletters).

## Partenariat scientifique et recherche confiée

Partenariat scientifique et recherche confiée  
  
N/A

## Références bibliographiques

El Bahlouli (2024). \*L’impact pédagogique des agents conversationnels en éducation : revue de littérature scientifique\*. Disponible sur : https://doi.org/10.3917/lfa.226.0027  
Velkovska, Relieu (2020). \*Pourquoi ethnographier les interactions avec les agents conversationnels ?\*. Disponible sur : https://doi.org/10.3917/res.220.0009

ANNEXES[[51]](#footnote-2)

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  
- L’impact pédagogique des agents conversationnels en éducation : revue de littérature scientifique (2024) — El Bahlouli — https://doi.org/10.3917/lfa.226.0027

## Verrou technique rencontré

1. EXERCICE : Période comptable ou d’activité d’une entreprise. [↑](#footnote-ref-21)
2. XX : Peut désigner une numérotation, des chromosomes ou être utilisé comme marqueur générique. [↑](#footnote-ref-55)
3. CLIENT : Personne ou entreprise qui achète un produit ou service. [↑](#footnote-ref-11)
4. LOGO : Symbole graphique représentant une marque ou une entreprise. [↑](#footnote-ref-34)
5. SOMMAIRE : Résumé ou table des matières d’un document. [↑](#footnote-ref-48)
6. ELEMENTS : Parties ou composants d’un ensemble. [↑](#footnote-ref-18)
7. VALORIS : Peut être un nom propre ou une abréviation, contexte nécessaire pour préciser. [↑](#footnote-ref-54)
8. DU : Contraction de "de le", préposition française. [↑](#footnote-ref-16)
9. EL : Peut désigner "élément" ou être utilisé comme prénom ou abréviation. [↑](#footnote-ref-17)
10. ES : Peut signifier "Espagne" (code pays) ou "est" (verbe être). [↑](#footnote-ref-20)
11. RAG : Méthode d’IA combinant recherche et génération de texte (Retrieval-Augmented Generation). [↑](#footnote-ref-42)
12. PME : Petite ou Moyenne Entreprise, selon la taille et le chiffre d’affaires. [↑](#footnote-ref-40)
13. IA : Intelligence Artificielle, discipline qui vise à simuler l’intelligence humaine avec des machines. [↑](#footnote-ref-26)
14. NLP : Traitement automatique du langage naturel (Natural Language Processing). [↑](#footnote-ref-37)
15. PDF : Format de fichier pour documents numériques (Portable Document Format). [↑](#footnote-ref-39)
16. LLM : Modèle de langage de grande taille (Large Language Model) en intelligence artificielle. [↑](#footnote-ref-33)
17. UI : Interface utilisateur d’un logiciel ou d’une application (User Interface). [↑](#footnote-ref-52)
18. UX : Expérience utilisateur lors de l’utilisation d’un produit ou service (User Experience). [↑](#footnote-ref-53)
19. F1 : Peut désigner "Formule 1" ou une touche de clavier pour l’aide. [↑](#footnote-ref-22)
20. YY : Peut désigner une numérotation ou être utilisé comme marqueur générique. [↑](#footnote-ref-57)
21. ZZ : Peut désigner une numérotation ou être utilisé comme marqueur générique. [↑](#footnote-ref-58)
22. CA : Chiffre d’Affaires, montant total des ventes d’une entreprise. [↑](#footnote-ref-8)
23. CIR : Crédit d’Impôt Recherche, aide fiscale pour les dépenses de recherche et développement. [↑](#footnote-ref-10)
24. MOOC : Cours en ligne ouvert à tous (Massive Open Online Course). [↑](#footnote-ref-36)
25. SIGKDD : Conférence internationale sur la découverte de connaissances à partir de données (Special Interest Group on Knowledge Discovery and Data Mining). [↑](#footnote-ref-47)
26. TCAF : Peut être une abréviation spécifique, contexte nécessaire pour préciser. [↑](#footnote-ref-50)
27. GPU : Processeur graphique utilisé pour le calcul intensif, notamment en IA. [↑](#footnote-ref-25)
28. DB : Base de Données, ensemble organisé d’informations stockées. [↑](#footnote-ref-13)
29. SQL : Langage informatique pour gérer des bases de données relationnelles (Structured Query Language). [↑](#footnote-ref-49)
30. BI : Ensemble d’outils et méthodes pour analyser des données et aider à la décision (Business Intelligence). [↑](#footnote-ref-6)
31. RBAC : Contrôle d’accès basé sur les rôles (Role-Based Access Control). [↑](#footnote-ref-44)
32. ID : Identifiant unique pour distinguer un élément ou une personne. [↑](#footnote-ref-27)
33. BPI : Banque Publique d’Investissement, organisme de financement des entreprises en France. [↑](#footnote-ref-7)
34. JEI : Jeune Entreprise Innovante, statut fiscal avantageux pour les startups innovantes. [↑](#footnote-ref-30)
35. CII : Crédit d’Impôt Innovation, aide fiscale pour les dépenses d’innovation des PME. [↑](#footnote-ref-9)
36. DES : Pluriel de "de", utilisé devant un nom pluriel. [↑](#footnote-ref-15)
37. OP : Abréviation possible pour "opération", "opérateur" ou "original poster" selon le contexte. [↑](#footnote-ref-38)
38. RATIONS : Portions ou quantités allouées de nourriture ou de ressources. [↑](#footnote-ref-43)
39. AU : Abréviation possible pour "Année Universitaire" ou "Autorisation d’Utilisation" selon le contexte. [↑](#footnote-ref-4)
40. TITRE : Nom ou désignation d’un document, d’un poste ou d’un sujet. [↑](#footnote-ref-51)
41. LEWIS : Nom propre, probablement une personne ou un contributeur. [↑](#footnote-ref-32)
42. IZACARD : Nom propre, probablement une personne ou un contributeur. [↑](#footnote-ref-29)
43. SHI : Peut être un nom propre ou une abréviation selon le contexte. [↑](#footnote-ref-46)
44. YIN : Concept du yin dans la philosophie chinoise, ou nom propre. [↑](#footnote-ref-56)
45. LSTM : Type de réseau de neurones utilisé pour traiter des séquences (Long Short-Term Memory). [↑](#footnote-ref-35)
46. BAHLOULI : Nom propre, probablement une personne ou un contributeur. [↑](#footnote-ref-5)
47. RELIEU : Nom propre, probablement une personne ou un contributeur. [↑](#footnote-ref-45)
48. FAISS : Librairie pour la recherche rapide de similarités dans de grandes bases de données. [↑](#footnote-ref-23)
49. API : Interface permettant à différents logiciels de communiquer entre eux. [↑](#footnote-ref-3)
50. PPT : Fichier de présentation créé avec PowerPoint. [↑](#footnote-ref-41)
51. ANNEXES : Documents ou informations complémentaires ajoutés à un dossier ou un rapport. [↑](#footnote-ref-2)