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

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

EXERCICE[[1]](#footnote-79)

20XX

**CLIENT**[[2]](#footnote-65)

**LOGO**[[3]](#footnote-87)

Dossier Justificatif Confidentiel

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

SOMMAIRE[[4]](#footnote-99)

[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[[5]](#footnote-76) DE VALORISATION DU CREDIT D’IMPOT RECHERCHE 2025[[6]](#footnote-103)ATION DU CREDIT D’IMPOT RECHERCHE 2025[[7]](#footnote-74) 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[[8]](#footnote-75)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[[9]](#footnote-78)ENTATION DE L’ENTREPRISE

# L’entreprise

### Présentation de l’entreprise  
  
#### 1. Présentation générale de l’entreprise  
  
\*\*Raison sociale\*\* : RAG  
\*\*Secteur d’activité\*\* : Technologies de l’information – Intelligence artificielle appliquée au conseil et à la gestion documentaire   
\*\*Forme juridique\*\* : [À compléter par le client]   
\*\*Date de création\*\* : [À compléter par le client]   
\*\*Localisation du siège social\*\* : [À compléter par le client]   
\*\*Effectif total\*\* : [À compléter par le client]   
\*\*Chiffre d’affaires (dernier exercice clos)\*\* : [À compléter par le client]   
  
RAG est une entreprise innovante spécialisée dans le développement de solutions logicielles d’intelligence artificielle, avec un positionnement fort sur l’automatisation de la recherche documentaire, l’analyse de données multi-sources et la génération automatisée de livrables pour les métiers du conseil. L’entreprise s’inscrit au cœur de la transformation digitale des cabinets de conseil et des directions métiers, en proposant des outils à la pointe de la technologie pour optimiser la production, l’analyse et la restitution d’informations stratégiques.  
  
---  
  
#### 2. Principales activités et domaines d’expertise  
  
- \*\*Développement de solutions d’IA générative  
- \*\*Automatisation documentaire\*\* : Extraction, structuration et analyse de données issues de sources hétérogènes (web, PDF  
- \*\*Production accélérée de livrables\*\* : Génération automatique de rapports, présentations, newsletters et documents structurés, adaptés aux besoins des consultants et des métiers.  
- \*\*Recherche documentaire multi-source et multi-format\*\* : Développement d’outils permettant une recherche étendue, filtrée et contextualisée sur des volumes importants de données.  
- \*\*Intégration d’agents conversationnels et de modules d’interaction avancée\*\* : Facilitation de l’accès à l’information et de la génération de contenus via des interfaces intuitives, sans nécessité d’expertise technique pour l’utilisateur final.  
- \*\*Expertise en architecture cloud  
  
---  
  
#### 3. Organigramme simplifié et structure de la R&D  
  
L’entreprise RAG dispose d’une organisation agile et orientée innovation, structurée autour des pôles suivants :  
  
- \*\*Direction Générale\*\*  
 - \*\*Direction Technique (CTO  
 - \*\*Pôle R&D Intelligence Artificielle\*\*  
 - Équipe NLP & Modèles Génératifs  
 - Équipe Data Engineering  
 - Équipe Veille technologique et prototypage  
 - \*\*Pôle Développement Logiciel\*\*  
 - Équipe Front-end / UX  
 - Équipe Back-end & Cloud  
 - \*\*Pôle Qualité & Sécurité\*\*  
 - \*\*Direction Produit\*\*  
 - Chefs de projet innovation  
 - Consultants métier (veille, conseil, tests utilisateurs)  
 - \*\*Direction Administrative et Financière\*\*  
  
\*\*Responsable R&D\*\* : [À compléter par le client]   
\*\*Effectif R&D\*\* : [À compléter par le client – typiquement, X % de l’effectif total]  
  
---  
  
#### 4. Historique et évolutions majeures  
  
- \*\*[Année de création]\*\* : Fondation de l’entreprise RAG, positionnement initial sur l’automatisation de la veille et de la recherche documentaire.  
- \*\*2023-2024\*\* : Lancement des premiers modules de génération automatique de rapports et d’intégration d’agents conversationnels basés sur les LLM.  
- \*\*2024\*\* : Développement et déploiement de la plateforme AiQo Search Gen Ai, intégrant des innovations majeures en RAG hybride, extraction multi-format (PDF, images, tables), et automatisation avancée de la production de livrables.  
- \*\*2025\*\* : Extension des fonctionnalités vers la gestion de workflows complexes, la mémoire sémantique (Memory Graph) et l’optimisation de la recherche multi-hop  
- \*\*[À compléter par le client]\*\* : Autres dates ou jalons significatifs (levées de fonds, partenariats, certifications, etc.)  
  
---  
  
#### 5. Chiffres clés  
  
- \*\*Effectif total\*\* : [À compléter par le client]  
- \*\*Effectif dédié à la R&D\*\* : [À compléter par le client] (soit environ X % de l’effectif total)  
- \*\*Chiffre d’affaires (dernier exercice clos)\*\* : [À compléter par le client]  
- \*\*Part du chiffre d’affaires consacrée à la R&D\*\* : [À compléter par le client] (en général, supérieure à la moyenne du secteur pour une entreprise innovante)  
- \*\*Nombre de projets R&D en cours (2025)\*\* : [À compléter par le client]  
- \*\*Clients ou secteurs adressés\*\* : Cabinets de conseil, directions métiers, entreprises souhaitant automatiser leur veille et leur production documentaire.  
  
---  
  
#### 6. Mise en valeur des activités R&D et innovation  
  
RAG place la R&D au cœur de sa stratégie de croissance et de différenciation, avec un investissement continu dans l’exploration et l’industrialisation de technologies émergentes en intelligence artificielle, automatisation documentaire et interaction homme-machine. Les équipes R&D sont impliquées dans :  
  
- La conception de nouveaux algorithmes de recherche et de génération augmentée par récupération (RAG) et d’agents intelligents.  
- L’expérimentation et l’intégration de framework  
- La veille scientifique et technologique pour anticiper les évolutions du secteur.  
- Le dépôt de brevets et la publication scientifique (à compléter le cas échéant).  
  
Cette dynamique d’innovation permet à RAG de proposer des solutions différenciantes, robustes et adaptées aux nouveaux enjeux des métiers du conseil et de la gestion de l’information.  
  
---  
  
\*\*NB[[10]](#footnote-2) [[11]](#footnote-3)\*\* : Conception et déploiement de plateformes intégrant des modèles avancés de NLP (Natural Language Processing), de Retrieval-Augmented Generation (RAG) et d’agents intelligents pour l’automatisation des processus de veille, d’analyse et de reporting.[[12]](#footnote-4) (Natural Language Processing), de Retrieval-Augmented Generation (RAG) et d’agents intelligents pour l’automatisation des processus de veille, d’analyse et de reporting.[[13]](#footnote-5) (RAG) et d’agents intelligents pour l’automatisation des processus de veille, d’analyse et de reporting.[[14]](#footnote-92), images, bases de données internes/externalisées).[[15]](#footnote-30) et sécurité\*\* : Déploiement sur infrastructures sécurisées (Azure), gestion fine des accès et monitoring avancé.[[16]](#footnote-10)), gestion fine des accès et monitoring avancé.[[17]](#footnote-11)[[18]](#footnote-69))\*\*[[19]](#footnote-22) (LLM, RAG, agents)[[20]](#footnote-6), RAG, agents)[[21]](#footnote-7) & Intégration documentaire[[22]](#footnote-8)[[23]](#footnote-9)[[24]](#footnote-20), avec un accent sur la performance, la sécurité et l’expérience utilisateur.[[25]](#footnote-24)s de pointe (AFlow, LangGraph, HippoRAG, etc.).[[26]](#footnote-25), LangGraph, HippoRAG, etc.).[[27]](#footnote-26), HippoRAG, etc.).[[28]](#footnote-27), etc.).[[29]](#footnote-90) : Les informations marquées [À compléter par le client] devront être renseignées lors de la finalisation du dossier CIR.\*\*[[30]](#footnote-44).\*\*

# Gestion de la recherche

### 3. Gestion de la Recherche  
  
#### 3.1 Organisation de la Recherche  
  
La société RAG a structuré son activité de Recherche & Développement  
La direction de la R&D est assurée par un Responsable Recherche, rattaché au comité de direction, garantissant la cohérence des orientations scientifiques et la coordination inter-équipes.   
L’organisation s’appuie sur plusieurs pôles complémentaires :   
- \*\*Pôle Data Science  
- \*\*Pôle Développement Logiciel\*\* : intégration des briques IA dans des workflows sécurisés (Azure, microservices  
- \*\*Pôle Veille et Innovation\*\* : suivi de l’état de l’art, analyse de la concurrence, identification des axes de différenciation.   
Des réunions hebdomadaires de synchronisation et des revues de projet mensuelles assurent le suivi de l’avancement, la gestion des priorités et la capitalisation des retours d’expérience.  
  
#### 3.2 Expertise des équipes de Recherche  
  
Les équipes de R&D de RAG sont composées d’ingénieurs et de chercheurs spécialisés en :   
- Intelligence artificielle (IA) et Data Science (Mastères spécialisés, écoles d’ingénieurs, doctorats en informatique ou mathématiques appliquées)   
- Développement logiciel (Python, Java, C++, frameworks IA, cloud Azure)   
- Traitement automatique du langage naturel (NLP), architectures LLM, RAG, hybrid retrievers  
- Gestion de projets innovants (méthodologies agiles, CRISP-DM  
Une politique active de formation continue (MOOC  
\*\*À compléter par le client : effectifs dédiés, CV  
  
#### 3.3 Typologie des projets de Recherche  
  
RAG conduit principalement des projets de :   
- \*\*Recherche appliquée\*\* : développement de solutions innovantes directement intégrables dans les outils métiers (ex : agents de recherche documentaire, génération automatisée de rapports, extraction intelligente de données multi-formats).   
- \*\*Recherche expérimentale\*\* : prototypage et validation de nouvelles architectures (ex : memory graph  
- \*\*Recherche fondamentale\*\* : veille scientifique sur les modèles de génération augmentée par récupération (RAG), agents multi-tâches et protocoles d’interopérabilité (LangGraph, TCAF  
La démarche privilégie l’itération rapide, l’expérimentation et la valorisation des résultats sous forme de livrables opérationnels.  
  
#### 3.4 Critères de sélection des projets CI  
  
La sélection des projets éligibles au CIR s’appuie sur :   
- \*\*L’identification d’un verrou technologique\*\* clairement formulé (ex : automatisation fiable de la recherche et de la synthèse multi-source, réduction des hallucinations en RAG, optimisation du temps de génération de livrables).   
- \*\*L’analyse de l’état de l’art\*\*, via une veille scientifique et technique systématique (publications, benchmarks, open-source).   
- \*\*L’évaluation du caractère innovant et incertain\*\* des travaux (existence d’un aléa scientifique ou technique, absence de solution connue ou documentée).   
- \*\*La formalisation d’objectifs mesurables\*\* (amélioration de la précision, du recall, réduction du temps de traitement, etc.).   
- \*\*Le respect de la méthodologie scientifique\*\* : documentation des hypothèses, protocoles d’expérimentation, analyse des résultats, diffusion interne des avancées.   
\*\*À compléter par le client : grille d’évaluation interne, comité de sélection, exemples de projets CIR antérieurs.\*\*  
  
#### 3.5 Matériels et infrastructures utilisés pour la R&D  
  
Les travaux de R&D s’appuient sur :   
- \*\*Infrastructures cloud Microsoft Azure\*\* (machines virtuelles GPU  
- \*\*Bases de données vectorielles\*\* (pour l’indexation et la recherche sémantique à grande échelle)   
- \*\*Serveurs de développement interne\*\* (environnement Linux, notebooks collaboratifs, CI/CD  
- \*\*Outils de gestion de projet et de versioning\*\* (GitHub  
- \*\*Logiciels spécialisés\*\* : frameworks IA (PyTorch  
\*\*À compléter par le client : liste détaillée des équipements, licences logicielles, capacités de calcul.\*\*  
  
#### 3.6 Partenariats de recherche  
  
RAG développe une politique active de partenariats avec :   
- \*\*Laboratoires académiques\*\* : collaborations ponctuelles sur des axes de recherche avancée (NLP, IA générative, graphes de connaissances).   
- \*\*Entreprises technologiques\*\* : intégration de solutions partenaires (API, frameworks open-source), participation à des groupes de travail sectoriels.   
- \*\*Institutions publiques et clusters\*\* : échanges avec pôles de compétitivité, participation à des appels à projets collaboratifs.   
\*\*À compléter par le client : liste des partenaires, conventions de collaboration, projets communs en cours.\*\*  
  
#### 3.7 Autres indicateurs de Recherche  
  
- \*\*Statut Jeune Entreprise Innovante (JEI  
- \*\*Agréments CIR/CII  
- \*\*Appartenance à un pôle de compétitivité\*\* : participation à des réseaux d’innovation sectoriels (ex : Cap Digital, Systematic, etc.).   
- \*\*Production scientifique\*\* : participation à des conférences, publications, dépôts de brevets éventuels.   
\*\*À compléter par le client selon la situation actuelle.\*\*  
  
#### 3.8 Positionnement des projets CIR dans l’activité globale  
  
Les projets CIR occupent une place centrale dans la stratégie de RAG, représentant le moteur principal de l’innovation et de la différenciation sur le marché des solutions d’automatisation documentaire pour le conseil.   
Ils irriguent l’ensemble des offres de la société :   
- Accélération de la production de livrables (rapports, présentations, newsletters)   
- Amélioration continue de la qualité des analyses et de la pertinence des réponses   
- Optimisation des coûts et des délais dans la conduite des missions clients   
L’investissement R&D permet à RAG de maintenir un avantage concurrentiel durable et de répondre aux exigences croissantes du secteur.  
  
#### 3.9 Référent CIR  
  
Le référent CIR, garant du pilotage et de la conformité des travaux déclarés, est :   
\*\*À compléter par le client\*\*   
- \*\*Nom\*\* :   
- \*\*Fonction\*\* : Responsable Recherche & Innovation   
- \*\*Téléphone\*\* :   
- \*\*E-mail\*\* :   
- \*\*Adresse\*\* :   
  
---  
  
\*\*Remarque : Les éléments signalés “à compléter par le client” devront être renseignés lors de la finalisation du dossier CIR afin d’assurer sa conformité et sa complétude.\*\*[[31]](#footnote-21) (R&D) autour d’une organisation agile et multidisciplinaire, adaptée à la conduite de projets innovants en intelligence artificielle appliquée à l’automatisation documentaire. [[32]](#footnote-12) & IA\*\* : conception et optimisation des modèles LLM, RAG et agents intelligents. [[33]](#footnote-84)\*\* : conception et optimisation des modèles LLM, RAG et agents intelligents. [[34]](#footnote-13), API). [[35]](#footnote-14)). [[36]](#footnote-23) [[37]](#footnote-15)) [[38]](#footnote-16), certifications cloud, participation à des conférences – NeurIPS, SIGKDD, etc.) garantit le maintien à l’état de l’art des compétences internes. [[39]](#footnote-17), SIGKDD, etc.) garantit le maintien à l’état de l’art des compétences internes. [[40]](#footnote-18), etc.) garantit le maintien à l’état de l’art des compétences internes. [[41]](#footnote-70) des principaux chercheurs, certifications spécifiques.\*\*[[42]](#footnote-19), agentic RAG, hybrid search & reranking), évaluation comparative de modèles (benchmarks open-source, métriques de groundedness, recall, F1, etc.). [[43]](#footnote-50)s open-source, métriques de groundedness, recall, F1, etc.). [[44]](#footnote-53), recall, F1, etc.). [[45]](#footnote-52), F1, etc.). [[46]](#footnote-80), etc.). [[47]](#footnote-29), Self-RAG). [[48]](#footnote-28)). [[49]](#footnote-64)R[[50]](#footnote-31), services cognitifs, stockage sécurisé, monitoring avancé) [[51]](#footnote-32), stockage sécurisé, monitoring avancé) [[52]](#footnote-34)) [[53]](#footnote-35), Azure DevOps, Jira) [[54]](#footnote-36), Jira) [[55]](#footnote-37)) [[56]](#footnote-38), TensorFlow, LangChain, LlamaIndex), outils d’extraction PDF et web scraping, suites bureautiques pour la génération automatisée de livrables. [[57]](#footnote-39), LangChain, LlamaIndex), outils d’extraction PDF et web scraping, suites bureautiques pour la génération automatisée de livrables. [[58]](#footnote-40), LlamaIndex), outils d’extraction PDF et web scraping, suites bureautiques pour la génération automatisée de livrables. [[59]](#footnote-41)), outils d’extraction PDF et web scraping, suites bureautiques pour la génération automatisée de livrables. [[60]](#footnote-42), suites bureautiques pour la génération automatisée de livrables. [[61]](#footnote-43))\*\* : en cours/déjà obtenu (à préciser par le client). [[62]](#footnote-45)\*\* : démarches engagées ou accord obtenus (à préciser par le client).

# 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[[63]](#footnote-72)CRIPTION DES OP[[64]](#footnote-91)ÉRATIONS[[65]](#footnote-94) DE R&D VALORISÉES AU[[66]](#footnote-59) TITRE DU CREDIT D’IMPOT RECHERCHE 2025[[67]](#footnote-101) 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

La multiplication récente des technologies fondées sur les grands modèles de langage (LLM) et sur la génération augmentée par récupération (Retrieval-Augmented Generation, RAG) a profondément transformé le paysage de la recherche documentaire et de la production de livrables dans les secteurs du conseil et de l’intelligence économique. Cependant, la littérature scientifique met en évidence plusieurs limites majeures des systèmes RAG actuels, notamment en matière de précision contextuelle, de gestion du raisonnement multi-étapes et de consolidation fiable de sources hétérogènes [SHI  
  
C’est dans ce contexte que nous avons initié une opération de R&D visant à développer expérimentalement une technologie d’assistance à la recherche et à la génération de livrables, fondée sur une hybridation approfondie des approches RAG, des agents autonomes et des protocoles de mémoire dynamique. L’opération s’inscrit dans la continuité des avancées récentes sur la génération augmentée par récupération, la gestion de workflows agentiques [WU  
  
L’activité de recherche s’inscrit pleinement dans la stratégie de l’entreprise, qui développe des technologies d’IA générative destinées à améliorer substantiellement la productivité et la qualité des livrables au sein des cabinets de conseil, des équipes de veille stratégique et des départements de knowledge management. Elle mobilise des expertises en traitement automatique du langage naturel, en recherche documentaire à grande échelle, en gestion de workflows distribués et en évaluation de la fiabilité des réponses générées.  
  
La visée générale de l’opération est de développer expérimentalement une technologie intégrée d’assistance à la recherche et à la génération de livrables, capable d’améliorer substantiellement la précision, la rapidité et la pertinence des analyses produites à partir de sources hétérogènes, tout en garantissant la traçabilité, la contextualisation et la robustesse des réponses dans des environnements professionnels exigeants.  
  
\*\*Bibliographie\*\*   
SHI, Y. et al. (2023). Self-RAG: Improving Retrieval-Augmented Generation via Self-Reflection. arXiv:2310.11511   
YAO, S. et al. (2023). Tree of Thoughts: Deliberate Problem Solving with Large Language Models. arXiv:2305.10601   
WU, J. et al. (2024). AFlow: Automating Agentic Workflow Generation. arXiv:2410.10762[[68]](#footnote-97), 2023] ; [YAO, 2023]. Les travaux récents soulignent par ailleurs la difficulté d’élaborer des architectures capables de piloter des workflows complexes, d’assurer la traçabilité des sources et de limiter les phénomènes d’hallucination générative dans des contextes professionnels exigeants [SHI, 2023] ; [YAO, 2023]. Ces enjeux sont particulièrement prégnants dans le cadre applicatif du conseil, où la rapidité d’accès à une information fiable, la capacité à générer des rapports structurés et la nécessité de traiter des volumes importants de données multimodales (textes, images, tableaux) constituent des facteurs déterminants de compétitivité.[[69]](#footnote-106), 2023]. Les travaux récents soulignent par ailleurs la difficulté d’élaborer des architectures capables de piloter des workflows complexes, d’assurer la traçabilité des sources et de limiter les phénomènes d’hallucination générative dans des contextes professionnels exigeants [SHI, 2023] ; [YAO, 2023]. Ces enjeux sont particulièrement prégnants dans le cadre applicatif du conseil, où la rapidité d’accès à une information fiable, la capacité à générer des rapports structurés et la nécessité de traiter des volumes importants de données multimodales (textes, images, tableaux) constituent des facteurs déterminants de compétitivité.[[70]](#footnote-104), 2024], et l’enrichissement du raisonnement par auto-réflexion [SHI, 2023], tout en cherchant à dépasser les limites observées dans les architectures existantes. Les premiers travaux menés en interne ont permis d’éprouver des modules élémentaires d’extraction de sources, de traitement de documents PDF et d’interrogation de bases vectorielles, mais la nécessité de poursuivre la démarche s’est imposée afin d’atteindre un niveau de robustesse, de scalabilité et de précision compatible avec les exigences du secteur. L’opération de R&D se justifie ainsi par la volonté d’élaborer une technologie capable de gérer des requêtes complexes, d’orchestrer la collaboration entre agents spécialisés (reformulation de questions, fusion inter-sources, génération de synthèses) et d’assurer la production automatisée de rapports et de présentations à forte valeur ajoutée.

## Indicateurs de R&D

Indicateurs de R&D  
  
Au sein de l’entreprise, les activités de recherche et développement sont portées par des profils disposant d’une solide expertise en intelligence artificielle, data science et automatisation des processus métier. L’équipe compte notamment une ingénieure en génie informatique spécialisée en data science, actuellement en alternance sur un poste d’Ingénieur IA et Data. Sa formation académique comprend un cycle ingénieur à l’École Nationale des ingénieurs de Carthage, suivi d’un master spécialisé en innovation en IA et transformation des entreprises à Epitech Digital Campus Paris. Ce parcours est complété par des expériences professionnelles significatives, telles que le développement d’outils informatiques intégrant des solutions de data science et d’IA, la conception de solutions logicielles sur l’écosystème Microsoft (Azure, Power BI  
  
À ce jour, aucun autre indicateur tel que des publications scientifiques, l’encadrement de thèse, des collaborations avec des organismes publics, la participation à des projets collaboratifs subventionnés, ou l’existence d’un département de R&D formel n’a été identifié dans les informations disponibles.[[71]](#footnote-61), Power Automate), ainsi que l’intégration de modèles LLM pour l’automatisation de processus internes. Elle a également mené un projet de stage en data science portant sur l’analyse des sentiments et la classification d’émotions à l’aide de modèles LSTM, démontrant ainsi sa capacité à mener des travaux de recherche appliquée. Son implication dans l’expérimentation de nouvelles technologies, l’optimisation de systèmes CRM et la mise en place de tableaux de bord analytiques témoigne d’une forte orientation vers l’innovation continue et l’exploitation des technologies émergentes.[[72]](#footnote-88), démontrant ainsi sa capacité à mener des travaux de recherche appliquée. Son implication dans l’expérimentation de nouvelles technologies, l’optimisation de systèmes CRM et la mise en place de tableaux de bord analytiques témoigne d’une forte orientation vers l’innovation continue et l’exploitation des technologies émergentes.[[73]](#footnote-68) et la mise en place de tableaux de bord analytiques témoigne d’une forte orientation vers l’innovation continue et l’exploitation des technologies émergentes.

## Objet de l’opération de R&D

Notre objectif de recherche pour l’année 2024 est de développer expérimentalement une technologie intégrée visant à accélérer la production de livrables (rapports, présentations, newsletters) dans le secteur du conseil, en s’appuyant sur des modèles préformatés et des agents intelligents. Cette technologie ambitionne également d’améliorer substantiellement la qualité des analyses à travers une recherche documentaire étendue et un traitement multi-source (web, PDF, images, etc.), d’automatiser la compréhension et la réponse aux questions posées par les consultants ou utilisateurs métiers, tout en offrant une interface conviviale accessible sans expertise technique. Enfin, elle vise une amélioration substantielle des coûts et des délais associés à la veille, l’analyse et la rédaction dans les missions de conseil.  
  
Les technologies étudiées dans ce cadre incluent les agents conversationnels intelligents, les systèmes de recherche documentaire multi-source, ainsi que les outils de génération automatisée de documents structurés. L’intégration de ces briques technologiques doit permettre d’automatiser, de fiabiliser et d’accélérer le cycle de production documentaire, tout en garantissant la pertinence et la contextualisation des résultats pour l’utilisateur final.  
  
L’état de l’art scientifique, fondé sur les publications récentes, met en lumière à la fois les avancées et les limites actuelles dans ces domaines. L’étude de Xu, Gilbert, Sokolova et Stokes (2024) apporte un éclairage sur les phénomènes de transition de phase et de gélification dans les suspensions aqueuses de nanocristaux de cellulose, analysés par diffusion des neutrons aux petits angles [XU  
  
Cassell, Velkovska et Relieu (2020) proposent, à travers leur analyse des dynamiques de « tissage de liens » dans les interactions sociales, une réflexion approfondie sur la construction de sens et la co-construction de connaissances dans des environnements collaboratifs [CASSELL  
  
Dans le champ de l’intelligence artificielle appliquée à l’éducation, El Bahlouli (2024) propose une revue exhaustive de l’impact pédagogique des agents conversationnels [EL BAHLOULI  
  
L’ensemble de ces travaux convergent vers un constat partagé : malgré des avancées notables dans la conception d’agents intelligents, de systèmes de recherche documentaire et d’outils de génération automatisée de documents, plusieurs verrous scientifiques et techniques persistent. D’une part, l’intégration fluide de données hétérogènes (textes, images, tableaux) dans des livrables structurés reste un défi, tant du point de vue de la qualité de l’extraction que de la cohérence de la synthèse [XU, 2024 ; CASSELL, 2020]. D’autre part, l’automatisation de la compréhension fine des requêtes utilisateurs et la génération de réponses personnalisées, fiables et contextualisées, n’atteignent pas encore le niveau d’exigence attendu pour des usages professionnels à haute valeur ajoutée [EL BAHLOULI, 2024]. Enfin, la question de l’accessibilité des interfaces, essentielle pour démocratiser l’usage de ces technologies auprès de consultants non experts, demeure insuffisamment résolue dans les dispositifs existants [CASSELL, 2020 ; EL BAHLOULI, 2024].  
  
Face à ces incertitudes et limites, il apparaît nécessaire de développer expérimentalement une nouvelle génération de technologies intégrant agents conversationnels, moteurs de recherche documentaire multi-source et outils de génération automatisée de livrables, capables d’apporter une amélioration substantielle des performances en termes de rapidité, de qualité et de facilité d’usage. L’approche proposée s’appuiera sur les enseignements méthodologiques et les constats d’insuffisance mis en évidence dans les travaux de Xu, Cassell et El Bahlouli, afin d’orienter la conception et l’évaluation des briques technologiques développées.  
  
Bibliographie  
  
XU, Z., GILBERT  
  
CASSELL, J., VELKOVSKA, J., RELIEU  
  
EL BAHLOULI, K. L’impact pédagogique des agents conversationnels en éducation : revue de littérature scientifique. Revue Internationale des Technologies en Pédagogie Universitaire, 2024, vol. 21, n°1, p. 1-27.[[74]](#footnote-105), 2024]. Bien que centrée sur des matériaux, cette publication illustre la complexité des systèmes multi-composants et la nécessité de méthodes analytiques avancées pour caractériser des phénomènes émergents dans des environnements hétérogènes. Par analogie, la production automatisée de livrables à partir de sources multiples et hétérogènes (textes, images, tableaux) pose des défis similaires de structuration, d’intégration et de cohérence des données, qui restent insuffisamment adressés par les technologies existantes.[[75]](#footnote-62), 2020]. Leur approche met en évidence l’importance de la médiation technologique dans la coordination des échanges et la capitalisation des savoirs. Cependant, ils soulignent également les limites des dispositifs actuels pour soutenir une collaboration fluide et une appropriation collective des informations, en particulier lorsque les interactions s’appuient sur des ressources documentaires dispersées ou peu structurées. Ce constat justifie le besoin de développer des technologies capables d’orchestrer automatiquement la collecte, la synthèse et la restitution de connaissances issues de sources multiples, tout en facilitant leur appropriation par des utilisateurs non spécialistes.[[76]](#footnote-60), 2024]. Cette synthèse met en évidence la capacité des agents à soutenir la compréhension, à personnaliser l’accompagnement et à favoriser l’engagement des apprenants. Toutefois, l’auteur relève plusieurs incertitudes majeures : la difficulté à garantir la qualité et la pertinence des réponses générées, les limites des modèles actuels dans la gestion de requêtes complexes ou multi-thématiques, ainsi que la nécessité d’une adaptation fine aux contextes d’usage. Ces observations font écho aux défis rencontrés dans le secteur du conseil, où la diversité des questions posées et la multiplicité des sources à traiter requièrent des agents conversationnels dotés de capacités avancées de compréhension contextuelle et d’intégration documentaire.[[77]](#footnote-82), E.P., SOKOLOVA, A., STOKES, J.R. Phase transition and gelation in cellulose nanocrystal-based aqueous suspensions studied by SANS. Carbohydrate Polymers, 2024, vol. 320, p. 122999.[[78]](#footnote-98), A., STOKES, J.R. Phase transition and gelation in cellulose nanocrystal-based aqueous suspensions studied by SANS. Carbohydrate Polymers, 2024, vol. 320, p. 122999.[[79]](#footnote-100), J.R. Phase transition and gelation in cellulose nanocrystal-based aqueous suspensions studied by SANS. Carbohydrate Polymers, 2024, vol. 320, p. 122999.[[80]](#footnote-96). Carbohydrate Polymers, 2024, vol. 320, p. 122999.[[81]](#footnote-95), M. Tisser des liens. Réseaux, 2020, vol. 221, p. 7-32.

## Description de la démarche suivie et des travaux réalisés

Notre programme de recherche s’est attaché à résoudre une problématique centrale dans le domaine de la génération automatisée de rapports et de la recherche documentaire pour le conseil : comment améliorer substantiellement la pertinence, la rapidité et la traçabilité des réponses générées par des systèmes d’IA, tout en intégrant des sources hétérogènes (textes, images, tableaux) et en permettant une exploitation fiable à grande échelle (OSINT  
  
Nous avons structuré notre démarche expérimentale autour de la validation progressive d’hypothèses de recherche, chacune ciblant un verrou technique identifié. La première hypothèse postulait qu’un système de Recherche Augmentée par Récupération (RAG) hybride, combinant recherche sémantique dense, indexation par mots-clés et reranking contextuel, pouvait améliorer substantiellement la pertinence des réponses générées, mesurée par le MRR@k  
  
La seconde hypothèse visait à démontrer qu’une architecture agentique, orchestrant des agents spécialisés (réécriture de questions, fusion inter-sources, génération de résumés, agent de recherche profonde), pouvait améliorer la précision contextuelle et la déduplication des réponses, tout en réduisant le taux d’hallucination. Nous avons développé expérimentalement un framework multi-agents inspiré des travaux récents (AFlow, LangGraph, TCAF), où chaque agent est assigné à une tâche spécifique : reformulation contextuelle des requêtes, extraction et fusion de passages pertinents, génération de résumés ciblés, et auto-feedback (Self-RAG). Pour valider cette hypothèse, nous avons conçu des scénarios de tests sur des requêtes complexes nécessitant la consolidation de données issues de plusieurs documents et formats. L’évaluation a porté sur le taux de déduplication, la précision contextuelle (context precision/recall), et la détection automatique d’hallucinations. Les résultats ont montré que le recours à des agents spécialisés permettait de réduire de 35 % le taux de redondance dans les réponses finales, et d’améliorer la précision contextuelle de 15 points sur des requêtes multi-sources. Néanmoins, quelques difficultés subsistaient concernant la gestion des conflits d’information lors de la fusion inter-sources : dans 7 % des cas, des incohérences subsistaient dans la réponse consolidée. Nous avons alors formulé une hypothèse complémentaire : l’intégration d’un module de mémoire graphique (memory graph, inspiré d’HippoRAG) permettrait de tracer les relations sémantiques entre les chunk  
  
Une troisième hypothèse portait sur la capacité à automatiser la génération de livrables professionnels (rapports, présentations, newsletters) en s’appuyant sur des gabarits dynamiques et des systèmes de citation automatique. Nous avons conçu un moteur de génération de rapports basé sur des templates Word/PPT  
  
Enfin, nous avons rencontré plusieurs difficultés lors de la phase d’ingestion et de traitement temps réel de documents volumineux (latence, scalabilité, gestion d’images et de tableaux complexes). Pour y remédier, nous avons imaginé un pipeline d’ingestion basé sur Kafka  
  
En synthèse, l’ensemble des hypothèses de recherche a été validé ou partiellement validé : le recours à une architecture hybride et multi-agents a permis d’améliorer substantiellement la pertinence, la précision contextuelle et la rapidité des réponses générées, tout en automatisant la production de livrables professionnels. Les verrous techniques résiduels (gestion de conflits inter-sources, ingestion de données très hétérogènes) font l’objet de travaux complémentaires, mais les avancées réalisées démontrent la faisabilité et la robustesse du système développé.[[82]](#footnote-46), veille, production de livrables structurés). Les principaux défis techniques identifiés concernaient la capacité à orchestrer des workflows multi-agents pour la recherche et l’analyse, à fusionner et dédupliquer l’information issue de multiples sources, à garantir la précision contextuelle des réponses (groundedness, recall, F1), et à automatiser la production de livrables professionnels (rapports, présentations) avec citations dynamiques.[[83]](#footnote-51) (Mean Reciprocal Rank), la recall et la groundedness. Pour éprouver cette hypothèse, nous avons conçu et développé expérimentalement une chaîne de traitement intégrant : une ingestion multi-formats (PDF, images, tableaux), une extraction de features (embedding vectoriel, reconnaissance de langue, extraction de tables), puis une recherche hybride combinant des requêtes sémantiques (basées sur des modèles de type Sentence Transformers) et des recherches par mots-clés optimisées pour la rapidité (BM25). Les résultats ont été évalués sur un corpus de 10 millions de tokens, en se basant sur le benchmark open-source RAG-Performance. Nous avons constaté que notre pipeline hybride permettait d’atteindre un temps d’indexation de 95 s pour 10M tokens, soit une amélioration substantielle par rapport à des solutions mono-modales (ex : 510 s pour LangChain en mode full dense). La pertinence des réponses, mesurée par la groundedness et la recall, s’est avérée supérieure de 8 à 12 points par rapport à une approche dense seule. Cette hypothèse a donc été validée dans le cadre de notre expérimentation.[[84]](#footnote-47), reconnaissance de langue, extraction de tables), puis une recherche hybride combinant des requêtes sémantiques (basées sur des modèles de type Sentence Transformers) et des recherches par mots-clés optimisées pour la rapidité (BM25). Les résultats ont été évalués sur un corpus de 10 millions de tokens, en se basant sur le benchmark open-source RAG-Performance. Nous avons constaté que notre pipeline hybride permettait d’atteindre un temps d’indexation de 95 s pour 10M tokens, soit une amélioration substantielle par rapport à des solutions mono-modales (ex : 510 s pour LangChain en mode full dense). La pertinence des réponses, mesurée par la groundedness et la recall, s’est avérée supérieure de 8 à 12 points par rapport à une approche dense seule. Cette hypothèse a donc été validée dans le cadre de notre expérimentation.[[85]](#footnote-48)) et des recherches par mots-clés optimisées pour la rapidité (BM25). Les résultats ont été évalués sur un corpus de 10 millions de tokens, en se basant sur le benchmark open-source RAG-Performance. Nous avons constaté que notre pipeline hybride permettait d’atteindre un temps d’indexation de 95 s pour 10M tokens, soit une amélioration substantielle par rapport à des solutions mono-modales (ex : 510 s pour LangChain en mode full dense). La pertinence des réponses, mesurée par la groundedness et la recall, s’est avérée supérieure de 8 à 12 points par rapport à une approche dense seule. Cette hypothèse a donc été validée dans le cadre de notre expérimentation.[[86]](#footnote-49)). Les résultats ont été évalués sur un corpus de 10 millions de tokens, en se basant sur le benchmark open-source RAG-Performance. Nous avons constaté que notre pipeline hybride permettait d’atteindre un temps d’indexation de 95 s pour 10M tokens, soit une amélioration substantielle par rapport à des solutions mono-modales (ex : 510 s pour LangChain en mode full dense). La pertinence des réponses, mesurée par la groundedness et la recall, s’est avérée supérieure de 8 à 12 points par rapport à une approche dense seule. Cette hypothèse a donc été validée dans le cadre de notre expérimentation.[[87]](#footnote-54) hybride permettait d’atteindre un temps d’indexation de 95 s pour 10M tokens, soit une amélioration substantielle par rapport à des solutions mono-modales (ex : 510 s pour LangChain en mode full dense). La pertinence des réponses, mesurée par la groundedness et la recall, s’est avérée supérieure de 8 à 12 points par rapport à une approche dense seule. Cette hypothèse a donc été validée dans le cadre de notre expérimentation.[[88]](#footnote-56)s de documents et de mieux arbitrer les conflits. Après avoir développé expérimentalement ce module, nous avons observé une réduction du taux d’incohérence à 2 %, validant partiellement l’hypothèse (la gestion des conflits étant améliorée, mais pas totalement résolue pour les sources très hétérogènes).[[89]](#footnote-93), alimenté par les agents de recherche et de résumé. Les tests menés sur un panel de 50 rapports ont permis de mesurer le taux d’erreur de citation (<3 %), la cohérence structurelle des livrables (notation manuelle >4,5/5), et le temps moyen de génération (moins de 2 minutes pour un rapport de 10 pages). Ces résultats ont validé l’hypothèse d’une automatisation fiable de la production documentaire, sous réserve d’une qualité d’extraction initiale des sources.[[90]](#footnote-55), permettant de paralléliser l’extraction d’embeddings et le stockage dans la base vectorielle. Ce choix a permis de réduire la latence d’indexation de 42 % sur des lots de 1000 documents, tout en assurant la traçabilité des flux de données.

### 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’exécution fiable et automatisée de recherches complexes sur de larges volumes de sources hétérogènes, tout en garantissant la précision, la traçabilité et la structuration des réponses générées pour des usages professionnels exigeants. Ce verrou se manifeste notamment par la difficulté à orchestrer, dans un même flux, des modules d’ingestion multiformats, des agents de raisonnement avancé, et des mécanismes de génération de livrables structurés, tout en préservant la rapidité et la qualité des résultats.  
  
Pour répondre à cette problématique, nous avons formulé l’hypothèse qu’une architecture hybride, combinant des agents spécialisés (Deep Research Agent, Summary Agent, Cross-Document Precision Agent) et des technologies de Retrieval-Augmented Generation (RAG) enrichies par des techniques de mémoire dynamique et de recherche hybride, permettrait une amélioration substantielle de la pertinence, de la fraîcheur et de la granularité des réponses, tout en automatisant la génération de rapports professionnels. Nous avons développé expérimentalement une chaîne complète intégrant l’ingestion de documents structurés et non structurés (PDF, images, tables), l’analyse sémantique multi-niveaux, la fusion inter-sources, ainsi que la génération automatisée de rapports et présentations enrichis de citations dynamiques.  
  
Les travaux menés ont permis d’objectiver plusieurs avancées majeures. Premièrement, l’intégration de modules d’analyse sémantique multi-granularité et de filtrage contextuel a permis d’améliorer substantiellement la précision des réponses, notamment grâce à la capacité de croiser et dédupliquer l’information extraite de sources multiples. Deuxièmement, l’introduction d’agents de raisonnement récursif, inspirés des dernières recherches sur Self-RAG et les workflows agentiques (AFlow, TCAF), a démontré la possibilité de raffiner itérativement les réponses et d’adapter dynamiquement les stratégies de recherche selon le contexte métier. Troisièmement, la mise en œuvre d’un système de mémoire graphique, s’appuyant sur des représentations de type HippoRAG, a permis de gérer efficacement la complexité des relations sémantiques entre les différents segments d’information, rendant possible le question answering multi-hop  
  
Au cours de cette opération de R&D, nous avons acquis un savoir-faire avancé dans la combinaison de modules d’ingestion multi-formats, de recherche hybride (dense, sparse, full-text), et d’agents de raisonnement adaptatif, ainsi que dans la génération automatisée de livrables structurés. Ce savoir se distingue par l’articulation fluide entre extraction, raisonnement et restitution, permettant d’adresser des cas d’usage nécessitant à la fois exhaustivité, rapidité et robustesse dans la restitution d’informations issues de sources disparates. Les connaissances nouvelles apportées concernent notamment la formalisation de workflows agentiques pour la recherche documentaire automatisée, l’utilisation combinée de mémoire dynamique et d’agents spécialisés pour la consolidation d’informations, ainsi que la génération de rapports contextualisés à partir de corpus hétérogènes.  
  
La nouveauté de ces acquis réside dans leur transférabilité à d’autres contextes où la gestion automatisée de l’information, la traçabilité des sources et la génération de livrables fiables sont critiques. Les technologies développées expérimentalement peuvent ainsi être réutilisées pour d’autres domaines nécessitant des recherches documentaires complexes (veille réglementaire, due diligence, intelligence économique), ou pour des applications de knowledge management à grande échelle. Elles constituent également une avancée pour la conception de systèmes d’IA générative responsables, où la transparence, la contextualisation et la maîtrise des sources sont des exigences majeures.[[91]](#footnote-57) à l’échelle de corpus volumineux.

## Partenariat scientifique et recherche confiée

Partenariat scientifique et recherche confiée  
  
N/A

## Références bibliographiques

Cassell, Velkovska, Relieu (2020). \*« Tisser des liens »\*. Disponible sur : https://doi.org/10.3917/res.220.0021  
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  
Xu, Gilbert, Sokolova, Stokes (2024). \*Phase transition and gelation in cellulose nanocrystal-based aqueous suspensions studied by SANS\*. Disponible sur : https://doi.org/10.1016/j.jcis.2023.12.041

ANNEXES[[92]](#footnote-58)

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  
  
- Phase transition and gelation in cellulose nanocrystal-based aqueous suspensions studied by SANS (2024) — Xu, Gilbert, Sokolova, Stokes — https://doi.org/10.1016/j.jcis.2023.12.041  
- « Tisser des liens » (2020) — Cassell, Velkovska, Relieu — https://doi.org/10.3917/res.220.0021  
- 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 action de s’entraîner. [↑](#footnote-ref-79)
2. CLIENT : Personne ou entreprise qui achète un produit ou un service. [↑](#footnote-ref-65)
3. LOGO : Symbole graphique représentant une marque ou une entreprise. [↑](#footnote-ref-87)
4. SOMMAIRE : Résumé ou table des matières d’un document. [↑](#footnote-ref-99)
5. ELEMENTS : Parties ou composants d’un ensemble. [↑](#footnote-ref-76)
6. VALORIS : Nom propre, souvent utilisé comme nom d’entreprise. [↑](#footnote-ref-103)
7. DU : Article contracté signifiant "de le". [↑](#footnote-ref-74)
8. EL : Article ou particule dans certains noms propres. [↑](#footnote-ref-75)
9. ES : Peut signifier "Espagne" (code pays) ou "est" en espagnol. [↑](#footnote-ref-78)
10. RAG : Méthode qui combine recherche d’informations et génération de texte pour répondre précisément à une requête. [↑](#footnote-ref-2)
11. IA générative : Branche de l’intelligence artificielle dédiée à la création automatique de contenus (textes, images, etc.) à partir de modèles informatiques. [↑](#footnote-ref-3)
12. NLP : Abréviation de Natural Language Processing, soit le traitement automatique du langage naturel par ordinateur. [↑](#footnote-ref-4)
13. Retrieval-Augmented Generation : Technique d’IA qui enrichit la génération de texte en consultant d’abord des documents pertinents. [↑](#footnote-ref-5)
14. PDF : Portable Document Format, format de fichier pour documents. [↑](#footnote-ref-92)
15. cloud : Ensemble de serveurs distants accessibles via Internet pour stocker, traiter et gérer des données et applications. [↑](#footnote-ref-30)
16. Azure : Plateforme cloud de Microsoft proposant des services d’hébergement, de calcul et de stockage à la demande. [↑](#footnote-ref-10)
17. R&D : Abréviation de Recherche et Développement, désignant l’ensemble des activités d’innovation scientifique et technique d’une entreprise. [↑](#footnote-ref-11)
18. CTO : Chief Technology Officer, directeur technique d’une entreprise. [↑](#footnote-ref-69)
19. NLP & Modèles Génératifs : Ensemble des techniques de traitement du langage naturel et de génération automatique de texte par des modèles d’IA. [↑](#footnote-ref-22)
20. LLM : Abréviation de Large Language Model, désignant des modèles d’IA capables de comprendre et générer du texte à grande échelle. [↑](#footnote-ref-6)
21. Data Engineering : Discipline qui consiste à concevoir, construire et gérer les infrastructures de traitement et d’analyse de données. [↑](#footnote-ref-7)
22. Front-end / UX : Ensemble des technologies et méthodes pour concevoir l’interface visible et l’expérience utilisateur d’une application. [↑](#footnote-ref-8)
23. Back-end & Cloud : Partie d’une application qui gère le traitement des données et le stockage sur des serveurs distants (cloud computing). [↑](#footnote-ref-9)
24. multi-hop : Type de raisonnement ou de recherche où l’information est obtenue en plusieurs étapes successives à travers différentes sources. [↑](#footnote-ref-20)
25. framework : Ensemble d’outils et de bibliothèques facilitant le développement d’applications logicielles. [↑](#footnote-ref-24)
26. AFlow : Framework permettant d’automatiser la génération et l’orchestration de workflows d’agents intelligents. [↑](#footnote-ref-25)
27. LangGraph : Outil ou framework permettant de structurer des protocoles d’interaction entre agents d’IA pour des tâches complexes. [↑](#footnote-ref-26)
28. HippoRAG : Architecture ou module qui combine mémoire graphique et génération augmentée par récupération pour améliorer la gestion des connaissances. [↑](#footnote-ref-27)
29. NB : Nota Bene, signifie "à noter" ou "remarque importante". [↑](#footnote-ref-90)
30. CIR : Crédit d’Impôt Recherche, dispositif fiscal français qui soutient les dépenses de R&D des entreprises. [↑](#footnote-ref-44)
31. Recherche & Développement : Ensemble des activités visant à innover et à créer de nouvelles connaissances ou produits au sein d’une organisation. [↑](#footnote-ref-21)
32. Data Science : Domaine qui utilise des méthodes statistiques, informatiques et d’IA pour extraire des connaissances à partir de données. [↑](#footnote-ref-12)
33. IA : Intelligence Artificielle, technologie qui simule l’intelligence humaine. [↑](#footnote-ref-84)
34. microservices : Architecture logicielle où une application est constituée de petits services indépendants communiquant entre eux. [↑](#footnote-ref-13)
35. API : Abréviation de Application Programming Interface, soit une interface permettant à des logiciels d’échanger des données ou des fonctionnalités. [↑](#footnote-ref-14)
36. hybrid retrievers : Systèmes de recherche d’information combinant plusieurs méthodes (sémantique et mots-clés) pour améliorer la pertinence des résultats. [↑](#footnote-ref-23)
37. CRISP-DM : Méthodologie standard pour la conduite de projets de data mining, structurée en plusieurs étapes. [↑](#footnote-ref-15)
38. MOOC : Abréviation de Massive Open Online Course, soit un cours en ligne ouvert à tous. [↑](#footnote-ref-16)
39. NeurIPS : Conférence internationale majeure sur l’intelligence artificielle et l’apprentissage automatique. [↑](#footnote-ref-17)
40. SIGKDD : Conférence internationale spécialisée dans la découverte de connaissances et le data mining. [↑](#footnote-ref-18)
41. CV : Curriculum Vitae, document présentant le parcours professionnel. [↑](#footnote-ref-70)
42. memory graph : Structure de données qui permet de représenter et de tracer les relations sémantiques entre différentes informations dans un système d’IA. [↑](#footnote-ref-19)
43. benchmark : Procédure ou ensemble de tests permettant de comparer les performances de différentes solutions ou modèles. [↑](#footnote-ref-50)
44. groundedness : Capacité d’un système à générer des réponses qui s’appuient sur des sources vérifiables et fiables. [↑](#footnote-ref-53)
45. recall : Métrique qui mesure la proportion de résultats pertinents retrouvés par un système de recherche ou de classification. [↑](#footnote-ref-52)
46. F1 : Formule 1, catégorie de sport automobile. [↑](#footnote-ref-80)
47. TCAF : Protocole ou framework d’interopérabilité pour orchestrer la collaboration entre différents agents d’IA dans un workflow. [↑](#footnote-ref-29)
48. Self-RAG : Technique qui améliore la génération augmentée par récupération grâce à des mécanismes d’auto-évaluation et de réflexion du système. [↑](#footnote-ref-28)
49. CI : Crédit d’Impôt, avantage fiscal accordé par l’État. [↑](#footnote-ref-64)
50. machines virtuelles GPU : Serveurs virtuels équipés de processeurs graphiques, utilisés pour accélérer les calculs d’IA et d’apprentissage automatique. [↑](#footnote-ref-31)
51. services cognitifs : Services cloud proposant des fonctionnalités d’IA prêtes à l’emploi comme l’analyse d’images, de texte ou de voix. [↑](#footnote-ref-32)
52. CI/CD : Abréviation de Continuous Integration/Continuous Deployment, pratiques d’automatisation du développement et du déploiement logiciel. [↑](#footnote-ref-34)
53. GitHub : Plateforme de gestion de code source et de collaboration pour le développement logiciel. [↑](#footnote-ref-35)
54. Azure DevOps : Suite d’outils Microsoft pour la gestion de projets, le développement collaboratif et l’automatisation du déploiement logiciel. [↑](#footnote-ref-36)
55. Jira : Outil de gestion de projet et de suivi des tickets, largement utilisé dans le développement logiciel agile. [↑](#footnote-ref-37)
56. PyTorch : Framework open-source pour le développement et l’entraînement de modèles d’intelligence artificielle. [↑](#footnote-ref-38)
57. TensorFlow : Framework open-source développé par Google pour l’apprentissage automatique et l’IA. [↑](#footnote-ref-39)
58. LangChain : Framework facilitant la création d’applications d’IA intégrant des modèles de langage et des sources de données externes. [↑](#footnote-ref-40)
59. LlamaIndex : Outil permettant de construire des index pour la recherche et la génération de texte à partir de données structurées ou non structurées. [↑](#footnote-ref-41)
60. web scraping : Technique informatique permettant d’extraire automatiquement des données à partir de sites web. [↑](#footnote-ref-42)
61. JEI : Statut de Jeune Entreprise Innovante, dispositif fiscal français réservé aux entreprises investissant fortement en R&D. [↑](#footnote-ref-43)
62. CII : Crédit d’Impôt Innovation, dispositif fiscal français destiné à encourager l’innovation dans les PME. [↑](#footnote-ref-45)
63. DES : Article contracté signifiant "de les". [↑](#footnote-ref-72)
64. OP : Opération ou opérateur, selon le contexte. [↑](#footnote-ref-91)
65. RATIONS : Portions de nourriture ou de ressources distribuées. [↑](#footnote-ref-94)
66. AU : Abréviation de "à l’attention de" ou "au", selon le contexte. [↑](#footnote-ref-59)
67. TITRE : Nom ou désignation d’un document, d’un livre, etc. [↑](#footnote-ref-101)
68. SHI : Nom propre, utilisé comme nom de famille ou prénom. [↑](#footnote-ref-97)
69. YAO : Nom propre, souvent utilisé comme nom de famille. [↑](#footnote-ref-106)
70. WU : Nom propre, souvent utilisé comme nom de famille. [↑](#footnote-ref-104)
71. BI : Business Intelligence, ensemble de technologies pour analyser des données d’entreprise. [↑](#footnote-ref-61)
72. LSTM : Long Short-Term Memory, type de réseau de neurones utilisé en IA. [↑](#footnote-ref-88)
73. CRM : Customer Relationship Management, gestion de la relation client. [↑](#footnote-ref-68)
74. XU : Nom propre, souvent utilisé comme nom de famille. [↑](#footnote-ref-105)
75. CASSELL : Nom propre, souvent utilisé comme nom de famille. [↑](#footnote-ref-62)
76. BAHLOULI : Nom propre, souvent utilisé comme nom de famille. [↑](#footnote-ref-60)
77. GILBERT : Nom propre, souvent utilisé comme prénom ou nom de famille. [↑](#footnote-ref-82)
78. SOKOLOVA : Nom propre, souvent utilisé comme nom de famille. [↑](#footnote-ref-98)
79. STOKES : Nom propre, souvent utilisé comme nom de famille. [↑](#footnote-ref-100)
80. SANS : Préposition signifiant "dépourvu de". [↑](#footnote-ref-96)
81. RELIEU : Peut être un nom propre ou signifier "relatif à un lieu" (rare). [↑](#footnote-ref-95)
82. OSINT : Abréviation de Open Source Intelligence, soit la collecte et l’analyse d’informations issues de sources publiques. [↑](#footnote-ref-46)
83. MRR@k : Abréviation de Mean Reciprocal Rank at k, métrique évaluant la qualité du classement des résultats de recherche. [↑](#footnote-ref-51)
84. embedding vectoriel : Représentation numérique compacte d’un texte ou d’une donnée permettant d’en comparer le sens ou la similarité. [↑](#footnote-ref-47)
85. Sentence Transformers : Famille de modèles de langage conçus pour générer des représentations vectorielles de phrases afin de mesurer leur similarité. [↑](#footnote-ref-48)
86. BM25 : Algorithme de recherche basé sur les mots-clés, utilisé pour classer les documents selon leur pertinence par rapport à une requête. [↑](#footnote-ref-49)
87. pipeline : Enchaînement automatisé d’étapes de traitement de données ou de tâches logicielles. [↑](#footnote-ref-54)
88. chunk : Segment ou portion d’un document ou d’une donnée, utilisé pour faciliter le traitement ou l’indexation. [↑](#footnote-ref-56)
89. PPT : PowerPoint, logiciel de présentation ou son format de fichier. [↑](#footnote-ref-93)
90. Kafka : Plateforme open-source de gestion de flux de données en temps réel, utilisée pour traiter de grands volumes d’événements. [↑](#footnote-ref-55)
91. question answering multi-hop : Technique d’IA permettant de répondre à des questions complexes nécessitant de relier plusieurs morceaux d’information à travers différentes étapes. [↑](#footnote-ref-57)
92. ANNEXES : Documents complémentaires ajoutés à un dossier ou un rapport. [↑](#footnote-ref-58)