

Laurianne DEPROST

**Master Traitement Automatique des Langues
parcours Ingénierie Linguistique**

RAPPORT D'ALTERNANCE

[Alternance effectuée du 21 septembre 2020 au 13 septembre 2021]

EDF

32 Avenue Pablo Picasso, 92000 Nanterre

**Assistants virtuels :
application et enrichissement de l'offre de service**

Tutrice entreprise : **Aleksandra LEWICKA**
Tutrice académique : **Iris ESHKOL-TARAVELLA**

Soutenu le 25 juin 2021
Université Paris Nanterre
200 Avenue de la République 92001 Nanterre cedex
Année Universitaire 2020 / 2021

Remerciements

Mes premiers remerciements vont aux enseignants du master TAL-IL pour leurs cours de qualité, leur pédagogie, leur bienveillance et l'adaptabilité dont ils ont fait preuve face aux conditions sanitaires actuelles. L'année fut tout de même très enrichissante tant sur le plan scolaire que sur le plan humain.

De plus, cette année n'aurait pu être aussi riche sans l'équipe du groupe Intelligence Artificielle. Elle m'a permis d'apprendre de toutes les différences que nous pouvions avoir en termes de formations ou bien de personnalités. De fait, l'ambiance bienveillante et la curiosité de chacun donnent à ce pôle une dynamique intellectuelle stimulante. Je remercie donc l'ensemble des membres du groupe IA : Adèle, Aleksandra, Aliénor, Joanne, Idriss, Mathilde, Monyrâth, Nathan, Nicolas, Patrice, Rahim, Paul, Sami, Stéphane, Yasmina.

Je ne peux m'empêcher de remercier particulièrement Aleksandra pour m'avoir si bien encadrée, avec professionnalisme et honnêteté, ainsi que pour ses nombreux encouragements empreints de bienveillance et de confiance.

Enfin, je remercie Aude, Aleksandra et Adèle pour m'avoir donné l'opportunité de vivre cette année intense, enrichissante humainement et professionnellement. J'ai beaucoup appris grâce à cette équipe cosmopolite et interdisciplinaire. Cette alternance m'a permis de m'épanouir dans un cadre bienveillant et très inspirant.

Table des matières

Introduction	4
1. Environnement de travail.....	5
1.1. Présentation de l'entreprise EDF	5
1.1.1. Raison d'être de l'entreprise	7
1.2. Présentation du groupe IA	8
1.2.1. Composition du groupe IA.....	10
1.2.2. L'écosystème autour du groupe IA.....	10
1.2.3. Les différents types de projets au sein du groupe IA	12
2. Présentation des projets réalisés	14
2.1. Introduction à l'offre de service SAVE.....	15
2.1.1. SAVE - Service d'Assistants Virtuels d'Entreprise	15
2.1.2. Solution Do You Dream Up.....	16
2.1.3. Tour d'horizon des chatbots EDF	17
2.2. Mise en œuvre de SAVE à travers un cas d'usage : le chatbot EDF Recrute. 18	
2.2.1. Accompagnement du métier	19
2.2.2. Amélioration continue du chatbot.....	20
2.3. Enrichissement de SAVE à travers un cas d'usage : le chatbot EDF Hydro . 25	
2.3.1. Naissance du projet chatbot EDF Hydro	25
2.3.2. Déroulement de la phase de cadrage	28
2.3.3. Automatisation de la constitution de la base de connaissances	32
2.3.4. Poursuite du travail.....	40
Conclusion.....	41
Annexes.....	43

Table des Figures

Figure 1: Mix énergétique EDF, chiffres au 31/12/2020	5
Figure 2: DTEO (simplifié)	8
Figure 3: Écosystème IA interne en 2020	11
Figure 4: Écosystème IA externe en 2020	12
Figure 5: Cycle de vie d'un projet en PMPG.....	12
Figure 6: Le cycle d'un projet au sein du groupe IA.....	13
Figure 7: Périmètre et état actuel de l'offre de service SAVE.....	15
Figure 8: Services proposés dans le cadre de la MCO de l'offre de service SAVE	16
Figure 9: Tableau de bord du BMS pour le chatbot de la direction RH (filtrage: 7j).....	17
Figure 10: Bénéfices d'un chatbot	17
Figure 11: Cartographie des chatbots en 2020.....	18
Figure 12: Pistes d'amélioration envisagées.....	22
Figure 13: Composition de l'équipe projet	22
Figure 14: Brainstorming des fonctionnalités de l'atelier Design	23
Figure 15: Teaser EDF Recrute	24
Figure 16: Onboarding type2 EDF Recrute	24
Figure 17: Chatbox EDF Recrute	24
Figure 18: Contexte actuel de la gestion des déchets industriels Hydro Centre	26
Figure 19: Validation du besoin.....	27
Figure 20: Constitution de l'équipe projet EDF Hydro	28
Figure 21: Planning découpé en quatre phases	28
Figure 22: Découpage de la phase de cadrage	28
Figure 24: Chaîne de traitement mise en place pour EDF Hydro.....	30
Figure 25: Organisation des données pour un déchet simple	31
Figure 26: Organisation des données pour une famille de déchet	31
Figure 27: Extrait d'une fiche déchet au format pdf.....	33
Figure 28: Lecture par la machine de la fiche Fig.27	33
Figure 29: Informations à extraire à partir d'un exemple d'une fiche déchet	34
Figure 31: Lecture de la rubrique « Nature du déchet » par la machine (cf. Fig.29).....	35
Figure 32: Extraction de la rubrique « Nature du déchet » (cf. Fig.29).....	35
Figure 33: Lecture par la machine	36
Figure 34: Rubrique « Filières de traitement » d'une fiche pdf.....	36
Figure 35: Répartition des informations au sein des colonnes type du fichier d'import	37
Figure 36: Extrait du data frame final pour le déchet Verre (non alimentaire)	37
Figure 37: Visualisation du même déchet côté interface utilisateur	38
Figure 38: Exemples de groupes de formulation de type 1	38
Figure 39: Extrait du data frame final avec ajout des groupes de formulations	39
Figure 40: Affichage des groupes de formulations dans le BMS	40

Introduction

Dans le cadre de mon Master 2 de Traitement Automatique des Langues (TAL), j'ai été recrutée en alternance au sein du groupe Intelligence Artificielle de la Direction des Services Informatiques et Télécoms d'EDF (DSIT). Par mes compétences en TAL, j'ai eu pour mission d'appuyer certains projets d'IA. C'est pourquoi, j'ai été affectée au pôle Solutions Éditeurs qui gère notamment les assistants virtuels du Groupe. J'ai alors découvert au cours de cette année, le monde de ces agents logiciels qui m'était jusqu'ici encore inconnu.

Les assistants virtuels font partie d'une offre de service d'entreprise, plus communément appelée SAVE, dont la particularité sera développée plus loin dans ce rapport. De la même manière, nous présenterons l'équipe, ses missions ainsi que l'écosystème dans lequel j'ai eu la chance d'évoluer au cours de cette année.

Les diverses missions que j'ai pu effectuer à travers les assistants virtuels durant mon alternance m'ont amenée à m'interroger sur deux points fondamentaux : Comment faciliter le travail d'un bot manager et, à travers des techniques utilisées en TAL, comment trouver des solutions complémentaires à des solutions éditeurs préalablement instaurées dans le cadre de l'offre de service SAVE ?

Parmi les différentes activités que j'ai eu à mener dans le cadre de mon travail, j'ai choisi de présenter dans ce rapport, deux chantiers majeurs :

Le premier permet d'expliquer en quoi consiste l'offre de service d'assistants virtuels d'EDF pour toutes ses directions métiers. Je présenterai cette partie à travers l'exemple du chatbot du site de recrutement EDF.

Le second décrit l'enrichissement de cette offre de service par un développement spécifique que j'ai pu réaliser moi-même. Ce module de TAL vient compléter l'offre déjà proposée par la DSIT. J'exposerai mon travail à travers l'exemple d'un nouveau chatbot mis en place pour la Direction EDF Hydro.

Je conclurai le rapport par un bilan de mon expérience au sein de l'entreprise EDF mais aussi des compétences acquises durant cette année enrichissante.

1. Environnement de travail

Tout d'abord, nous présenterons l'entreprise EDF, sa mentalité et ses engagements, puis nous mettrons en avant la composition du groupe IA dans lequel nous évoluons avant de mettre en exergue l'écosystème gravitant autour de ce dernier.

1.1. Présentation de l'entreprise EDF

Le 8 avril 1946 demeure une date importante en France car elle correspond au vote d'une loi qui établit la nationalisation des sociétés privées de production, de transport et de distribution de l'électricité. Dans ce contexte d'après-guerre, un Établissement à Caractère Industriel et Commercial (EPIC) voit le jour : Électricité de France (EDF). En 2004, suite à la volonté européenne d'ouvrir le marché énergétique français à la hauteur de 70%, EDF devient une Société Anonyme à conseil d'administration (SA). Elle ouvre son capital et entre en bourse en 2005. Ce changement de statut juridique lui permet de varier ses activités. Ainsi, aujourd'hui le groupe d'Électricité de France reste un des leaders mondiaux dans le secteur de l'énergie (premier fournisseur d'énergie mondiale jusqu'en 2017), il compte environ 165 200 salariés (recensés fin 2020) et sa production énergétique est dominée par le nucléaire (77%). Toutefois il engage désormais 37% de ses investissements bruts dans le développement des énergies renouvelables (cf. mix énergétique, Fig.1). Avec 90% de sa production non émettrice de CO₂, le groupe EDF prouve qu'il est le champion de la croissance bas carbone. Bien que deuxième au classement mondial, EDF demeure le premier producteur et fournisseur d'électricité en France. Il est également présent en tant que producteur en Europe (Italie, Allemagne, Belgique, Grande-Bretagne) et hors de l'Europe (Brésil, Mexique, Chili, Canada, Etats-Unis).

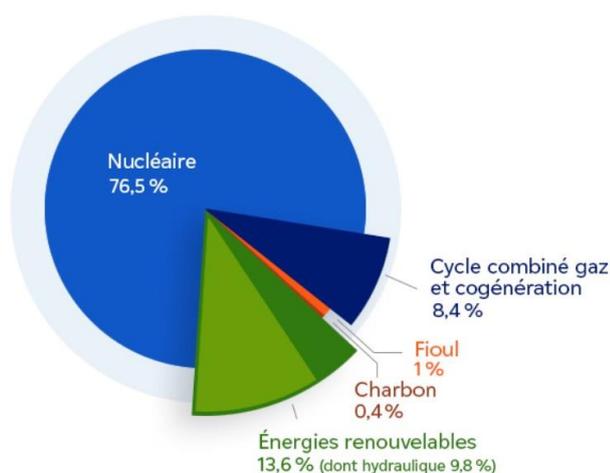


Figure 1: Mix énergétique EDF, chiffres au 31/12/2020

Hormis l'activité de production d'énergie, EDF fait de l'innovation un levier majeur de sa stratégie. Depuis 25 ans, l'entreprise investit dans la R&D pour préparer et inventer l'avenir de l'énergie. Trois plans stratégiques sont alors nés de ce nouveau challenge : le plan solaire, le plan stockage électrique et le plan mobilité électrique. Les trois projets visent donc à augmenter les capacités de stockage électrique dans le monde mais aussi répondre à une triple ambition : devenir le premier fournisseur en électricité pour véhicules, le premier exploitant de réseau de bornes électriques dans toute l'Europe, et enfin devenir le leader européen du « smart charging¹».

La poursuite de ces objectifs ne pourrait être possible sans la coordination d'un ensemble d'acteurs pariant sur la complémentarité de leurs activités. De ce fait, il est important de rappeler ces différents partenaires qui œuvrent au transport et à la distribution de l'énergie produite par EDF. Le réseau français d'acheminement de l'énergie électrique est organisé en 2 niveaux : le réseau de transport et le réseau de distribution. RTE est le gestionnaire du réseau de transport électrique en France, assurant le transport de l'électricité depuis les centres de production jusqu'aux grands sites industriels et aux réseaux de distribution qui prennent le relais vers les consommateurs. L'objectif de ce **transporteur** est d'assurer la continuité de l'alimentation en énergie électrique sur l'intégralité du territoire Français, tout en respectant les contraintes techniques, économiques et les conditions météorologiques de chaque région. Le réseau de distribution est quant à lui géré par la filiale ENEDIS qui assure l'acheminement de l'électricité chez le client final. A ce titre, ses responsabilités consistent à exploiter, gérer et entretenir le réseau de transport par lequel transite l'électricité. Ce **distributeur** s'occupe de ce fait du raccordement des foyers des abonnés, de la mise en service lors d'une souscription à un contrat, de dépannage ou intervention sur le réseau électrique. Enfin, nous comptons aussi les fournisseurs qui sont chargés de vendre des contrats d'énergie (gaz et électricité). Par ailleurs, l'activité de négoce d'énergie est réalisée par EDF Trading qui a pour mission de négocier les combustibles sur les marchés de gros (ou commerce inter-entreprises ou négoce) pour optimiser les demandes en quantité d'énergie et anticiper les risques.

Ainsi, l'entreprise EDF est issue d'une longue histoire et continue de s'adapter aux enjeux actuels tout en préparant l'avenir de l'énergie. Face à l'urgence climatique, et forte des valeurs de service public, EDF affirme vouloir s'engager pour cet avenir énergétique juste, innovant et durable. C'est notamment à travers la stratégie du groupe « CAP 2030 » que nous pouvons illustrer l'ancrage de cette entreprise dans une dynamique résolument innovante. Ce plan stratégique CAP 2030 affiche trois priorités : « Proximité clients », « Production bas carbone » et « Développement international ». Le premier vise à créer de nouvelles solutions compétitives décentralisées, de nouveaux services énergétiques personnalisés et des réseaux intelligents. En ce sens, le Groupe propose à ses clients de réaliser des économies d'énergie par des conseils, suivi de consommation, travaux de rénovation, accompagnement dans leur transition énergétique... Par exemple, en février

¹ technologies visant à optimiser la charge voire la décharge d'un véhicule électrique, en gérant la puissance de recharge du véhicule de façon efficace, flexible et économique (<https://edf.fr/strategie-cap-2030/mobilite-electrique/smart-charging>).

2019, l'entreprise lance sa nouvelle marque IZI by EDF pour offrir aux clients Particuliers et Professionnels une large gamme de services. Désormais, le Groupe commercialise des services de proximité, en particulier des petits travaux à la demande, des prestations de rénovation, de la fourniture et pose d'équipements, ainsi que des produits et services de ses filiales du Groupe EDF et de partenaires choisis. Le second axe majeur de cette stratégie « CAP 2030 » vise à rééquilibrer le mix de production en accélérant le développement des énergies renouvelables d'ici à 2030, tout en garantissant la sûreté et la performance du nucléaire existant et du Nouveau Nucléaire. Dans le cadre du troisième axe de son plan stratégique, l'entreprise envisage de tripler la part du business réalisé à l'international d'ici à 2030. En ce sens, elle souhaite se déployer sur de nouvelles géographies en développant ses solutions bas carbone dans les pays en croissance, tout en confortant ses positions en Europe. La finalité de ces stratégies actuellement menées par le Groupe est d'atteindre la neutralité totale en carbone d'ici à 2050.

1.1.1. Raison d'être de l'entreprise

La crise sanitaire actuelle a permis d'attirer plus largement l'attention de tous sur une autre crise, moins immédiate mais tout aussi périlleuse : le changement climatique, le défi de notre siècle. La crise que nous traversons accentue le rôle essentiel de l'énergie, et en particulier de l'électricité, pour bien vivre : pour nous nourrir, pour travailler à distance, pour rester connectés avec nos proches, pour être soigné. La responsabilité d'EDF est de rendre cette énergie accessible à tous, sur tous les territoires, en proposant des solutions pour moins et mieux consommer. C'est pourquoi, le 7 mai 2020, le groupe EDF a adopté sa Raison d'être, fruit de la contribution des salariés lors des dialogues Parlons Energies :

« Notre raison d'être, c'est de construire un avenir énergétique neutre en CO₂, conciliant préservation de la planète, bien-être et développement, grâce à l'électricité et à des solutions et services innovants »²

En effet, depuis la loi Pacte de mai 2019, les entreprises sont invitées à intégrer dans leur statut la manière dont elles entendent jouer un rôle dans la société, au-delà de leur seule activité économique.

Pourtant, il n'est pas rare de voir cette raison d'être résumée à la **neutralité carbone**. Un enjeu essentiel, bien sûr, puisque le Groupe s'est engagé à être neutre en CO₂ à l'horizon 2050. Toutefois, le Groupe souligne trois autres enjeux tout aussi importants. **La préservation de la planète**, qui comprend les actions et engagements du Groupe en matière de biodiversité, de gestion de l'eau, des sols, du traitement de ses déchets, radioactifs y compris. Le **bien-être**, de ses salariés comme de toutes ses parties prenantes : un enjeu clé, qui intègre les actions du Groupe pour la santé et la sécurité de tous, l'égalité et la diversité, ainsi que la lutte contre la précarité énergétique. Et enfin, l'enjeu de « **développement** ». On

² La raison d'être EDF est accompagnée d'un manifeste (cf. Annexes).

parle ici du développement de nouveaux projets bas carbone, du développement des territoires, sans oublier le développement des filières industrielles.

Une raison d’être c’est un engagement que nous prenons vis-à-vis de la société. Ce n’est pas un slogan publicitaire. Il s’agit d’un combat dans lequel chaque salarié doit se retrouver. Face à l’urgence environnementale actuelle, il me semble indispensable de repenser l’avenir. Cela passe notamment par notre manière de consommer, repenser notre modèle énergétique pour qu’il soit respectueux de la planète. De plus, je reste persuadée que la protection de l’environnement est un enjeu qui peut rassembler les français. Je suis fière, en tant que Femme salariée d’EDF, de participer à la transformation de demain.

1.2. Présentation du groupe IA

La Direction de la Transformation et Efficacité Opérationnelle (DTEO) est l’une des directions du Groupe EDF, elle délivre des services aux entités du Groupe et comme son nom l’indique, a pour ambition d’accompagner les métiers dans la réussite de leur transformation. Les services concernent les domaines suivants : achats du Groupe EDF, immobilier, informatique et télécoms, tertiaire (comptabilité, RH, ...), conseil.

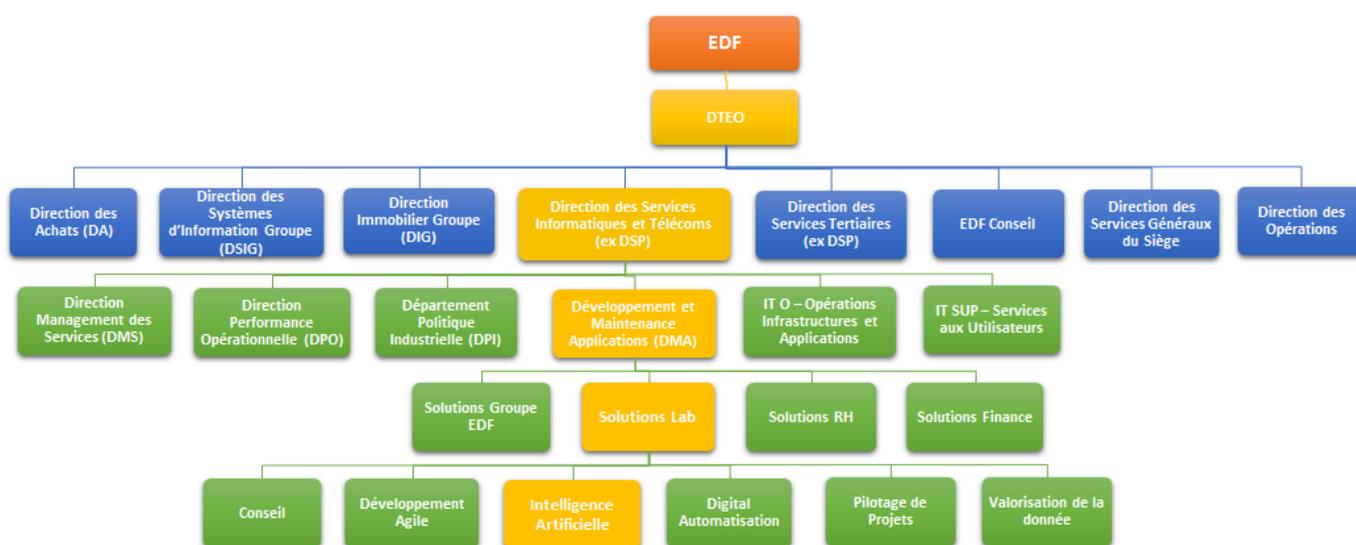


Figure 2: DTEO (simplifié)

Le groupe Intelligence Artificielle (aussi appelé GIA) appartient à cette direction et plus précisément à l’entité de Développement et Maintenance Applicative (DMA) (cf. Fig.2). Cette dernière développe pour les différentes directions métiers des outils informatiques tout en garantissant leur maintien en conditions opérationnelles et ce dans un modèle industrialisé. Le département Solutions Lab est en charge du développement de nouvelles solutions innovantes nécessaires à la transformation digitale et de leur industrialisation en vue d’un déploiement large. Ce point fait notre spécificité par rapport aux autres entités chez EDF.

Plusieurs groupes constituent le département Solutions Lab :

- Conseil qui a pour mission d'accompagner les métiers grâce à une importante connaissance du SI et d'autres domaines d'expertise.
- Pilotage de Projets dont le collectif des chefs de projets orchestre la mise en œuvre optimale de solutions SI grâce aux différentes méthodes de pilotage de projets (cycle en V, Agile, Saas, ...).
- Développement Agile apporte son expertise dans le développement et l'agilité sur des notions de mobilité, blockchain et tout ce qui touche au flux opérationnel (*workflow*).
- Digital Automatisation a pour objectif d'accompagner, d'outiller, de faciliter et automatiser l'industrialisation de solutions IA.
- Valorisation de la Donnée s'occupe de préparer les données, les croiser, les « faire parler » et élaborer des restitutions apportant de la valeur aux métiers. Il regroupe donc des activités de big data, d'ingénierie de données et de data visualisation.
- **Intelligence Artificielle** : créé en avril 2017, le pôle IA a connu une forte croissance jusqu'à devenir un groupe à part entière. Actuellement dirigé par Sami MALLEK, le groupe tend à concevoir et réaliser des outils dits d'intelligence artificielle. Le contexte économique actuel imposant la baisse des coûts, ces outils peuvent notamment aider à diminuer les frais de maintenance ou bien à réduire les tâches répétitives à faible valeur ajoutée en les automatisant.

Intelligence Artificielle

Il est crucial de rappeler ce qu'on appelle « Intelligence Artificielle ». En effet, il s'agit d'une technologie si vaste et révolutionnaire qu'il est difficile d'en donner une définition précise. Toutefois, le dictionnaire en ligne du Centre National de Ressources Textuelles et Lexicales (CNTRL) définit l'IA comme la « recherche de moyens susceptibles de doter les systèmes informatiques de capacités intellectuelles comparables à celles des êtres humains » (*La Recherche*, janv. 1979). En ce sens, on peut considérer qu'il s'agit d'une branche du domaine informatique dont le but est de créer des machines capables d'effectuer des tâches nécessitant traditionnellement une intelligence humaine. Les ordinateurs peuvent désormais, comme nous, humains, apprendre, analyser, raisonner, percevoir et même dialoguer. Cette discipline convoque aussi différents domaines tels que les mathématiques, l'informatique, les sciences du langage, les sciences cognitives, ainsi que différentes techniques telles que les ontologies, l'apprentissage automatique, l'apprentissage profond, etc. Nous retrouvons cette interdisciplinarité dans la composition de l'équipe ainsi que dans les projets menés par le groupe. De fait, les solutions développées par le groupe IA viennent en support afin d'augmenter l'intelligence des salariés, voire la remplacer totalement dans le cadre de certaines tâches répétitives et fastidieuses. Nous pouvons citer comme exemple la simplification du travail des métiers dans leur saisie de dossiers de rapports d'incidents ou encore le tri intelligent de mails.

1.2.1. Composition du groupe IA

Le groupe IA intervient pour l'ensemble des directions de l'entreprise, pour tout type de besoins en proposant des offres complètes clés en main, qui vont de l'émergence du besoin à l'élaboration des solutions, leur intégration et à leur maintien en production. Étant donné que toutes les entités font face à des tâches récurrentes et fastidieuses, les solutions mises en place par le groupe IA offrent un renfort, un gain de temps, simplifiant les activités.

Le groupe IA est composé de trois pôles : Solutions IA Open Source, Solutions IA Editeur ainsi que le pôle RPA (Automatisation Robotisée des Processus). Le premier réalise des solutions sur mesure pour répondre aux besoins métiers en s'appuyant sur des technologies innovantes. Les data scientists sont capables de traiter des données numériques, textuelles ou encore d'images pour en extraire de la valeur utile aux métiers leur permettant ainsi de gagner en efficacité opérationnelle. Le second pôle met en œuvre des solutions basées sur l'intelligence artificielle tout en s'appuyant sur les produits du marché, avec un hébergement soit sur les datacenters d'EDF soit en cloud. En grande majorité le groupe travaille avec des start-up françaises, souvent leaders dans leur domaine. L'équipe assure l'accompagnement des métiers tout au long de la vie de l'application ainsi que le lien avec l'exploitant et le fournisseur. Enfin, le pôle RPA réalise des solutions d'automatisation des processus robotisés qui permettent de reproduire automatiquement l'enchaînement de gestes SI ou métiers avec interface et connexion aux applications concernées. Le robot fonctionne seul sur un serveur dédié.

L'équipe du groupe IA est donc multiculturelle et interdisciplinaire. Elle regroupe 15 personnes dont 11 internes et 4 externes (consultants). La spécificité et la pluralité de l'équipe repose sur la transversalité des profils. En effet, le groupe se compose de data scientists, chefs de projet, product owners, des pilotes d'applications (POS), architectes, développeurs et une chargée de communication. A cela s'ajoute le responsable du groupe qui manage les équipes et détecte les opportunités auprès des métiers. Le groupe mutualise donc ces différentes compétences pour proposer une offre complète aux différents métiers et filiales de l'entreprise.

1.2.2. L'écosystème autour du groupe IA

1.2.2.1. Interne

Le groupe IA fonctionne main dans la main avec différents acteurs internes au Groupe EDF.

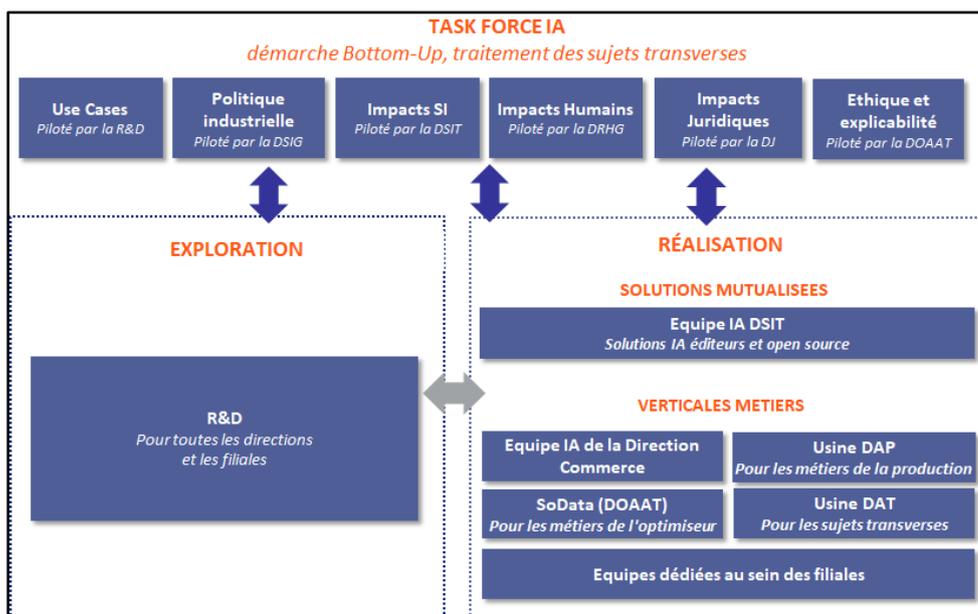


Figure 3: Écosystème IA interne en 2020

Créée en 2017 à la demande du COMEX, la Task force (force d'intervention) IA est un réseau important car il a pour objectif de mutualiser et créer des synergies entre les besoins et les projets IA au sein des différentes entités du Groupe EDF. A travers l'Ambition IA Groupe, le COMEX a pris la décision fin 2019 de faire de l'IA le moteur de la transformation numérique du Groupe. Pour ce faire, 6 groupes de travail ont été créés (cf. Fig.3).

Plus largement, il existe de nombreuses entités au sein d'EDF qui travaillent sur l'IA. Chacune apporte une vision spécifique, ce qui définit un périmètre large et complémentaire de l'IA. Ainsi, la R&D réalise en amont un travail de veille sur les nouveaux outils et autres innovations. Cette entité mène des études et intervient sur la phase exploratoire pour toutes les entités et filiales du Groupe. Elle ne réalise pas d'industrialisation à la différence de la DSIT dont les solutions développées ont pour enjeu de devenir des solutions d'entreprises. Nous comptons également des équipes IA au sein des verticales métiers qui travaillent pour des secteurs d'activités spécifiques. Par exemple, le groupe IA de la Direction Commerce (DSIN) travaille davantage sur des problématiques orientées clients particuliers, alors que les Usines Data Analytics s'adressent à des métiers spécifiques. Nous nous devons aussi d'évoquer le rôle de La Fabrik, une usine digitale créée en 2017 pour regrouper tous les services et offres numériques de la DSIT basées sur des solutions innovantes. Son rôle consiste à accompagner tous les métiers du Groupe dans leur transformation digitale et la construction de nouveaux services industrialisés et mutualisés.

1.2.2.2. Externe

EDF intervient aussi à l'extérieur. En effet, le Groupe travaille avec le secteur académique via la mise en place de partenariats avec de grandes écoles et universités mais

également avec des organismes de recherche tels que les laboratoires INRIA et le CNRS. L'entreprise intervient aussi au niveau institutionnel, notamment à travers la participation à des dispositifs gouvernementaux tels que celui des « Certificats d'Economie d'Énergie » (CEE) qui vise à réduire la consommation énergétique des français. Dans le cadre de ce dispositif, l'entreprise souhaite aider les particuliers à financer leurs travaux de rénovation énergétique. Enfin, elle travaille aussi bien avec des start-up françaises qu'avec les GAFAs, avec de grands groupes notamment au travers d'associations telle IMA ou TOSIT, ainsi que les administrations qui échangent sur leurs réalisations en IA.



Figure 4: Écosystème IA externe en 2020

1.2.3. Les différents types de projets au sein du groupe IA

1.2.3.1. La gestion de projet au sein d'EDF

Le déroulement d'un projet au sein de la DSIT répond à une Politique de Management de Projet du Groupe EDF (la PMPG). La présente politique s'applique aux différentes entités du Groupe. Elle est prescriptive pour tous les projets ou programmes à forts enjeux, dont la décision d'investissement relève du Comité des Engagements du Comité Exécutif du Groupe (CECEG). De ce fait, elle s'applique aussi bien à un projet du domaine nucléaire que pour le développement d'une application IA à fort enjeu. Elle constitue également un guide pour les autres projets de moindre envergure car elle décrit le cycle de vie d'un projet. Ce dernier est découpé en 5 grandes phases, séparées par des jalons de décision. La politique précise également les responsabilités du client (MOA), du chef de projet (MOE) et de son équipe.



Figure 5: Cycle de vie d'un projet en PMPG

1.2.3.2. La gestion de projet au sein du groupe IA

Etant donné que les projets de petite envergure ne sont pas soumis à la PMPG, le groupe IA décline ses projets à partir de la méthode Agile. Cette gestion de projet nécessite une forte implication de la part du métier tout en privilégiant les individus et leurs interactions avant les processus et les outils mais implique aussi de savoir accepter les changements plutôt que de s'attacher à rester conforme aux plans. La démarche Agile est incrémentale et itérative. Elle se découpe en sprints (périodes de courte durée) qui sont en quelque sorte des mini-projets.

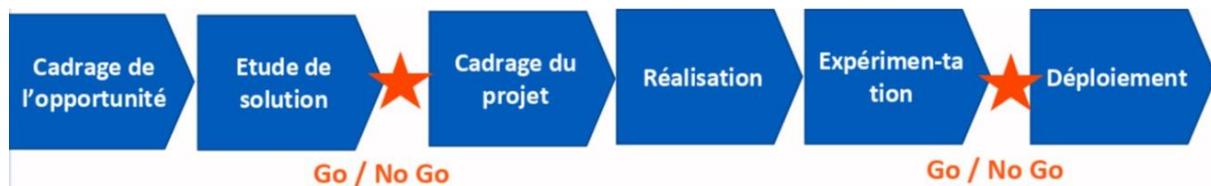


Figure 6: Le cycle d'un projet au sein du groupe IA

La première phase est celle du cadrage de l'opportunité. Elle consiste à comprendre le besoin Métier et à s'assurer de la pertinence des enjeux face à la stratégie du groupe. La seconde phase consiste à envisager une ou plusieurs solutions en réponse à la problématique métier. La phase cadrage du projet permet d'analyser plus profondément le besoin Métier et de disposer d'un plan de réalisation du projet. Durant cette phase, on précise la stratégie envisagée, on affine le périmètre, on distribue les rôles puis on anticipe l'intégration au SI. La phase suivante est la Réalisation, aussi appelée « build ». Cette phase se découpe en sprints de deux ou trois semaines. Durant la phase d'expérimentation, la maîtrise d'ouvrage (MOA) va tester la solution issue de la phase précédente. La MOA se réduit à un panel restreint d'utilisateurs et permet de faire remonter des anomalies et en apporter les corrections nécessaires. A l'issue de cette cinquième phase, le métier peut décider d'arrêter le projet si celui-ci juge que le prototype ne correspond pas à ses attentes ou pour toute autre raison. Si le métier donne son accord, le projet passe en phase d'industrialisation. Tout utilisateur final peut désormais accéder au produit déployé. La solution sera suivie par des outils de monitoring pour étudier son comportement et son adaptation dans le nouvel environnement de production. A la fin du cycle du projet, l'équipe projet passe le relais à l'équipe de maintenance qui va veiller à maintenir le produit en condition opérationnelle. Cette ultime étape s'appelle la MCO, aussi connue sous le nom de phase de « Run ».

1.2.3.3. La gestion de projet au sein du pôle Solutions IA Éditeurs

La gestion de projet du groupe IA et notamment le cycle de vie d'un projet en mode Agile ne peut pas s'appliquer de la même manière dans les trois pôles du groupe. En effet, les projets du pôle Solutions IA Éditeurs reposent sur des solutions du marché. Les différentes phases du cycle de vie d'un projet en méthode Agile que nous venons d'introduire sont donc

réajustées pour s'adapter aux projets du pôle Éditeurs. Toutefois, beaucoup s'inspirent des principes de la méthode Agile. En effet, nous retrouvons régulièrement la notion de sprints et d'étapes itératives. En revanche, nous ne retrouvons pas certaines notions clés de la méthode Agile telles que la réunion de mêlée quotidienne aussi appelée « stand-up daily » qui permet aux équipes de faire un point de coordination sur les tâches en cours et sur les difficultés rencontrées.

Quelques exemples de solutions du marché

L'équipe IA propose plusieurs solutions pour accompagner les métiers EDF dans leurs projets d'intelligence artificielle.

L'application AGATE permet d'optimiser le traitement des emails d'une boîte aux lettres générique (BAL) à fort volume de mails afin de les distribuer au bon interlocuteur avec une proposition de réponse. Cette solution offre diverses fonctionnalités telles que qualifier/catégoriser les mails, analyser des pièces jointes, proposer des réponses pertinentes et améliorer le suivi d'activité des BAL depuis des tableaux de bord. Ce service s'appuie sur la solution éditeur OWI, une start-up française.

Le projet Bot Analyzer est basé sur une solution développée par l'éditeur français YSEOP, l'un des leaders du marché NLG (génération automatique de texte). Leur outil permet, à partir des données structurées (tableaux excel, bases de données etc.) de rédiger ou aider à rédiger automatiquement des textes en langage naturel. La solution du marché est utilisée dans le cadre du projet Bot Analyzer pour le Centre de Services Partagés Comptabilité Conseil (CSP2D). L'outil permet de générer automatiquement des rapports financiers à partir de l'analyse de données financières extraites des applications métiers.

L'offre de service SAVE, basée sur la solution de l'éditeur français Do You Dream Up permet la mise en œuvre de chatbots et voicebots pour le Groupe EDF. Nous la détaillerons un peu plus tard dans ce rapport (cf. 2.1.1 SAVE – Service d'Assistants Virtuels).

2. Présentation des projets réalisés

Cette seconde partie sera consacrée à la présentation des projets menés au cours de cette alternance. Nous commencerons par présenter l'offre de service SAVE puis nous verrons un cas d'usage de son déploiement. Enfin nous verrons à travers le projet EDF Hydro qu'il peut être judicieux de l'enrichir, notamment par du développement interne.

2.1. Introduction à l'offre de service SAVE

Un assistant virtuel est un logiciel permettant d'interagir avec un utilisateur via une interface qui peut revêtir différentes formes : vocale, visuelle, textuelle. Un chatbot est un robot conversationnel qui dialogue par écrit avec les utilisateurs en langage naturel. Le robot « analyse » et « comprend » les intentions en langage naturel envoyées par les utilisateurs à travers sa base de connaissances. En effet, le cœur d'un chatbot repose sur un algorithme de compréhension du langage naturel (NLP). Cela consiste à utiliser un algorithme de classification automatique et reconnaissance d'entités nommées. Un des premiers traitements va alors consister à transformer la requête en texte. A cette étape, l'algorithme corrige l'orthographe, identifie les lemmes et repère les synonymes et hyperonymes. L'étape suivante consiste à effectuer un calcul de similarité sémantique entre les mots afin de comprendre le sens général de la phrase et ne pas interpréter chaque mot de manière isolée. Le programme découpe ensuite la demande de l'utilisateur en intentions et entités. Une fois le message traité, il cherche une correspondance entre la question posée et celles présentes dans la base de connaissances³ pour fournir une réponse à l'utilisateur.

2.1.1. SAVE - Service d'Assistants Virtuels d'Entreprise

Le groupe IA de la DSIT propose une plateforme mutualisée d'assistance virtuelle hébergée dans ses datacenters. Cette offre de service SAVE s'appuie sur une solution du marché : Do You Dream Up (DYDU). C'est donc à travers cette offre de service que 13 chatbots sont actuellement déployés pour différentes entités et filiales d'EDF. Par ce biais, les métiers EDF peuvent mettre à disposition de leurs utilisateurs un assistant virtuel avec lequel ils pourront dialoguer en langage naturel. Ce dernier permet de traiter des tâches à faible valeur ajoutée et d'apporter des informations à des clients internes et externes. Les chatbots permettent également de capitaliser des connaissances issues d'anciennes technologies et qui doivent être pérennisées pour de futurs métiers.

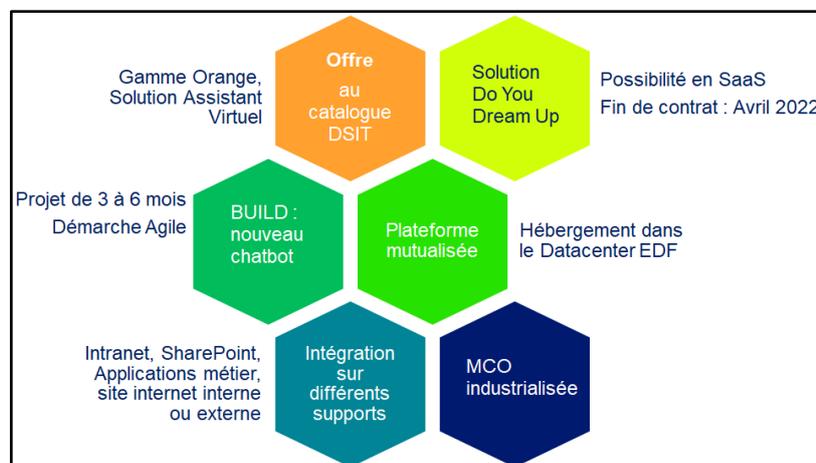


Figure 7: Périmètre et état actuel de l'offre de service SAVE

³ Une connaissance correspond au couple question/réponse et les formulations d'une question.

Le processus de mise en œuvre d'un chatbot comprend deux phases : une phase de Build au cours de laquelle le groupe IA accompagne le métier sur la construction et l'expérimentation de l'assistant virtuel et une phase de Run au cours de laquelle le groupe exploite et maintient le bot dans la durée. Ce dernier devient accessible par son intégration aux sites métiers (sites internet, intranet...). Dès lors qu'il y a industrialisation, le groupe assure le maintien en condition opérationnelle (MCO) du chatbot à travers plusieurs services.

Définition des rôles	Animation des métiers	Support technique et fonctionnel	Garantie de continuité de service	Montées de version régulières	Coordination	Relation éditeurs
<ul style="list-style-type: none"> Métier Bot Manager Product Owner <ul style="list-style-type: none"> Groupe IA POS/référent technique Référent fonctionnel/chef de projet Coordinatrice fournisseurs 	<ul style="list-style-type: none"> Réunions régulières Canal Teams partagé Information sur la vie de la plateforme Partage d'expérience 	<ul style="list-style-type: none"> Bonnes pratiques de gestion de connaissance Réponses techniques sur les besoins de paramétrage spécifique : vidéo, arbres de connaissance, carrousel etc. 	<ul style="list-style-type: none"> Disponibilité Incidents Sauvegardes Conduite applicative 	<ul style="list-style-type: none"> Applicatives Produit DYDU – trimestrielle Système Respect des exigences de sécurité et de pérennité des composants 	<ul style="list-style-type: none"> Audits de sécurité Revues semestrielle d'habilitations Travaux menés par les équipes des sites d'intégration 	<ul style="list-style-type: none"> Suivi contractuel rapproché

Figure 8: Services proposés dans le cadre de la MCO de l'offre de service SAVE

Nous porterons une attention particulière à l'accompagnement des métiers et à l'amélioration continue des performances des chatbots car ce sont des missions qui m'ont été confiées dans le cadre de mon alternance.

2.1.2. Solution Do You Dream Up

Do You Dream Up (DYDU) est une start-up française avec laquelle la DSIT travaille dans le cadre de son offre de service SAVE. Avec plus de 11 ans d'expérience, l'éditeur de logiciel est expert en intelligence conversationnelle, l'un des champs d'étude de l'IA. Leur outil permet de créer et gérer des robots conversationnels intelligents (chatbots, callbots, voicebots...) et fait la synthèse de 3 composants innovants : un algorithme NLP, un Bot Management System (BMS) pour gérer le paramétrage des solutions et une interface utilisateur pour interagir avec le chatbot.

- Le composant NLP va réaliser, entre autres, des calculs de distances entre la phrase de la requête de l'utilisateur et l'ensemble des connaissances présentes dans la base pour fournir la meilleure réponse possible à l'utilisateur. Un score inférieur à 80% engendre une demande de reformulation auprès de l'utilisateur tandis qu'un score inférieur à 50% annonce une incompréhension. Il est important de souligner que l'outil ne fait pas appel à l'apprentissage automatique. Ce choix repose sur une volonté de garder un contrôle sur les connaissances et une transparence sur le fonctionnement de l'outil.
- D'un point de vue interface Métier, l'éditeur propose au sein de sa console administrative de nombreuses fonctionnalités relativement intuitives. Il n'est donc pas nécessaire de savoir coder car le logiciel a été conçu pour être utilisable par tous, ce qui rend le métier autonome sur la gestion des connaissances. Par exemple, la base de connaissances d'un chatbot est directement injectée à partir d'un fichier Excel. Aussi,

certains outils disponibles sur la plateforme permettent l'enrichissement d'une base de connaissances, notamment par l'ajout de fonctionnalités plus complexes telles que des arbres de décision ou encore des groupes de formulation. Le BMS permet également de suivre la vie d'une solution, notamment à travers l'historique des conversations, des notifications d'alertes et à travers différents indicateurs statistiques de performance : avis utilisateur, conversations en échec, connaissances les plus sollicitées etc.

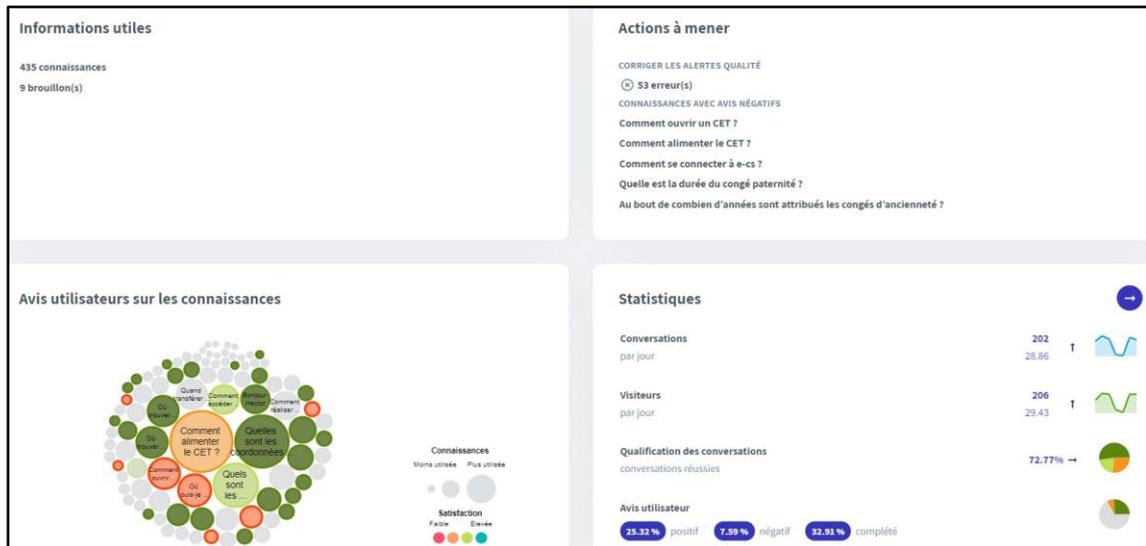


Figure 9: Tableau de bord du BMS pour le chatbot de la direction RH (filtrage: 7j)

2.1.3. Tour d'horizon des chatbots EDF

Les avantages du chatbot dans le monde de l'entreprise sont nombreux. Que son usage soit destiné aux clients ou aux collaborateurs, de nombreux bénéfices incontestables sont à souligner. Nous pouvons par exemple noter sa disponibilité inégalable qui permet aux utilisateurs d'obtenir des réponses instantanément, sans contrainte de disponibilité de leur interlocuteur ou encore de centraliser l'information par l'homogénéité des réponses apportées.



Figure 10: Bénéfices d'un chatbot

C'est donc tout naturellement que le Groupe EDF s'est lancé dans l'aventure des Assistants virtuels en 2011. Depuis, d'autres chatbots ont vu le jour et ont été déployés pour différentes entités et filiales d'EDF. Nous en comptons deux types : les chatbots dits « experts/sages » sont ceux qui capitalisent des connaissances, ils sont ouverts à très peu d'utilisateurs ; les chatbots aux ouvertures plus larges, à l'interne comme à l'externe et qui répondent à des sollicitations plutôt basiques, sur des thématiques récurrentes. Ces derniers capitalisent moins de connaissances mais connaissent une plus grande fréquentation de par leur plus large public.

Nous comptons actuellement 13 chatbots déployés dans le Groupe. Un nouveau est également en cours de développement (cf. 2.3.1 Naissance projet EDF Hydro). Parmi ces bots en production, seuls trois sont ouverts à l'externe : EDF Bot pour les clients particuliers d'EDF SA, EDF Recrute pour les candidats au recrutement ou salariés en mobilité interne et TCHAT'OA pour les producteurs d'énergie photovoltaïque partenaires du Groupe EDF.

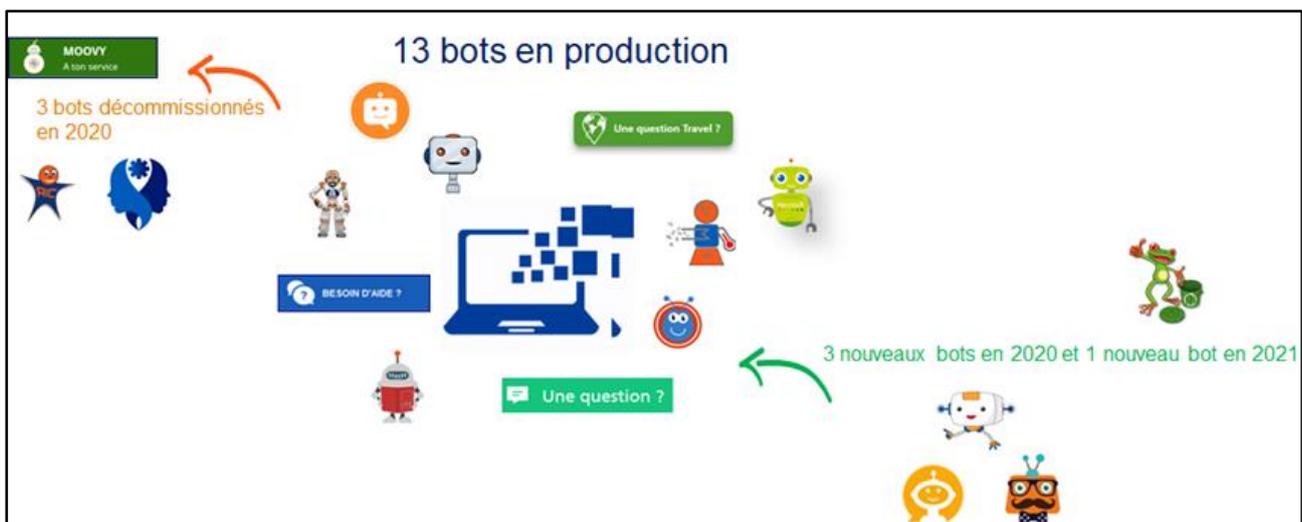


Figure 11: Cartographie des chatbots en 2020

2.2. Mise en œuvre de SAVE à travers un cas d'usage : le chatbot EDF Recrute

Le chatbot EDF Recrute a été mis en place par le groupe IA de la DSIT pour le service DRH Groupe. Il a été mis en production en juillet 2018 avec pour objectif de faciliter le parcours client sur le site de recrutement d'EDF. Les utilisateurs cibles de l'assistant virtuel sont de potentiels candidats au recrutement et donc des utilisateurs externes, voire internes au Groupe EDF. De ce fait, l'assistant virtuel permet à tout visiteur du site de poser des questions sur le recrutement au sein de Groupe EDF et ses filiales. Il est également sollicité pour des demandes d'assistance technique, type hotline. Par exemple, un utilisateur peut s'adresser au bot s'il rencontre des difficultés dans la constitution de son dossier de candidature ou dans l'insertion en ligne de ses pièces jointes. Enfin, l'assistant est également habilité à traiter des demandes administratives, comme par exemple fournir des attestations.

Le chatbot permet donc d'accompagner l'utilisateur dans ses demandes exprimées à l'écrit. A l'origine, le chatbot permettait également de donner l'état d'avancement de la candidature pour les personnes ayant postulé à des offres d'emploi, stage ou alternance. Toutefois, en raison d'un changement de solution, ce webservice attaché au chatbot a été abandonné.

Les bénéfices attendus suite à la mise de l'assistant virtuel sont principalement : un gain de productivité et une image moderne et digitale du Groupe EDF. En effet, la foire aux questions (FAQ) accessible auparavant sur le site ne suffisait plus à couvrir des demandes utilisateurs et les métiers RH se trouvaient régulièrement sur-sollicités par les candidats. La mise en place d'un chatbot a alors permis de contenir ce phénomène tout en modernisant l'image du Groupe très impliqué dans le numérique.

2.2.1. Accompagnement du métier

Dans le cadre de son offre de service, et particulièrement lors du maintien en conditions opérationnelles d'un assistant virtuel, le groupe IA assure l'accompagnement du métier tout au long de la vie de l'application et fait le lien avec l'exploitant et le fournisseur. Nous appelons « métier » ou encore « Bot Manager », la personne responsable du chatbot intégré dans son entité. Cet animateur de chatbot a un rôle essentiel dans la vie de l'assistant car il s'occupe de l'amélioration continue des performances de ce dernier mais promeut également son utilisation auprès des utilisateurs cibles.

Ainsi, dans le vocabulaire usuel de gestion de projet, nous pouvons dire que le groupe IA représente la Maîtrise d'Œuvre (MOE) car il va décider des moyens techniques à mettre en œuvre pour concevoir le produit conformément à ce qui a été défini par la Maîtrise d'Ouvrage (MOA), représentée par le Métier. Le rôle de chef de projet MOE est assuré par un membre du pôle Solutions IA Éditeurs. La prise en charge des 13 chatbots en production est aussi répartie entre les quatre membres de l'équipe : chefs de projet et POS - Pilote Opérationnel de Service. C'est pourquoi, au cours de mon alternance, on m'a confié la gestion de trois chatbots en production, dont EDF Recrute. Mon rôle est donc de représenter le client, c'est-à-dire le métier, et de prendre en charge les différents services que nous proposons dans le cadre de SAVE.

Dans ce nouveau rôle de chef de projet, j'ai été amenée à organiser des points de suivis avec les métiers. En effet, il est crucial de mettre en place ces réunions régulières pour suivre la vie d'un chatbot, identifier de potentiels points bloquants ou incidents mais aussi pour accompagner les évolutions et montées de versions de l'outil développé par DYDU. Plus simplement, ces points à fréquence mensuelle, bimensuelle ou hebdomadaire permettent d'accompagner les métiers dans leur rôle de Bot manager. Dans le cadre de cette animation métier avec Edf Recrute, nous avons instauré un point mensuel. Nous avons fait le choix de ce rythme selon l'activité du chatbot mais aussi car nous avons déjà d'autres chantiers en cours, dont nous parlerons plus loin dans ce rapport (cf. 2.2.2 Amélioration continue). Lors de ces réunions, j'ai eu l'occasion de conseiller le métier dans son rôle de Bot Manager, notamment au niveau de la gestion de la base de connaissances et dans l'explication de

certaines fonctionnalités ou paramétrages un peu plus techniques de l'outil. Par exemple, j'ai travaillé sur la constitution d'un court mode opératoire sur les arbres de connaissances. Ces moments m'ont permis de discuter des performances du bot après avoir mené de courtes analyses statistiques au sein du portail administratif (BMS) mais j'ai également pu aborder les chantiers évolutifs en cours, notamment en période de montée de versions. Enfin, ces temps d'échange sont des occasions pour discuter simplement de la vie du chatbot et d'en déduire des points bloquants et des pistes d'amélioration.

Toujours dans le cadre de cet accompagnement du métier, le pôle Éditeurs a mis en place un canal d'échange sur l'application TEAMS⁴. Ce canal de partage dédié aux Bots Managers a comme objectif de diffuser de l'information sur la vie de la plateforme mais il facilite également le partage d'expériences entre les bots managers. Par exemple, par ce biais, j'ai pu mettre à disposition de tous les bots managers mon mode opératoire sur les arbres de connaissances. Aussi, nous y avons signalé des anomalies identifiées ou encore des opérations de maintenance, lors de la nouvelle montée de version du BMS en avril 2021.

Par ailleurs, dans le cadre de l'offre de service SAVE, nous proposons également des formations générales aux fonctionnalités du BMS lors d'une arrivée d'un nouveau Bot Manager voire quand un métier ne se sent pas à l'aise avec l'outil. J'ai eu l'occasion de dispenser plusieurs fois ces formations, dont le niveau varie selon la maîtrise et l'expertise du métier dans ce domaine. J'ai notamment eu à dispenser cette formation à la personne en charge du chatbot EDF Recrute car cette dernière ne se sentait pas à l'aise avec l'outil. En effet, le BMS est un outil indispensable pour le Bot Manager. C'est pourquoi il est crucial de savoir tirer profit de ses fonctionnalités et paramétrages plus techniques afin de maximiser les performances d'un chatbot.

2.2.2. Amélioration continue du chatbot

2.2.2.1. Contexte

Un bot manager doit suivre régulièrement l'activité de son chatbot et veiller à l'amélioration continue de ses performances. Lors de ma prise en charge du chatbot EDF Recrute et par conséquent, des premiers points de suivis avec le Bot Manager, nous avons souvent évoqué l'état statique de l'assistant, en manque d'utilisateurs. Motivé à reprendre en main le chatbot, le métier m'a sollicitée pour l'accompagner et le guider dans ce processus.

Dans le but d'avoir une vision plus globale de la situation, j'ai commencé par me renseigner davantage sur des chatbots spécialisés dans le recrutement. J'ai alors réalisé une courte étude de la concurrence, aussi bien auprès de grands groupes que de start-up. J'ai rapidement remarqué que le chatbot EDF Recrute manquait un peu d'énergie par rapport aux assistants virtuels du marché. Après plusieurs échanges avec le bot manager et afin de

⁴ Application de communication collaborative propriétaire officiellement lancée par Microsoft en novembre 2016.

remédier à l'essoufflement d'EDF Recrute, nous avons décidé d'entamer une modernisation globale du chatbot. Suite au cadrage de ce projet, nous avons envisagé deux approches d'amélioration : l'aspect design et parcours utilisateurs mais aussi l'aspect technique via la mise à jour de la base de connaissances. Ces changements ont notamment pour objectif d'accroître la visibilité du chatbot, augmenter le nombre de réponses satisfaisantes fournies mais aussi d'améliorer son image et par conséquent celle du Groupe EDF.

2.2.2.2. Travail réalisé sur la base de connaissances

Pour redonner de la fraîcheur et du dynamisme au chatbot EDF Recrute, il est essentiel de ne pas se focaliser sur l'aspect design du chatbot mais de travailler également sur le cœur de l'outil, sa base de connaissances. En effet, l'aspect visuel d'un robot demeure important mais son côté technique l'est tout autant. Fournir une jolie solution virtuelle est insuffisant. Il est essentiel de veiller à ce que l'assistant soit pertinent dans les réponses qu'il donne aux utilisateurs. De plus, les chatbots fournis par l'éditeur DYDU n'ayant pas recours à l'apprentissage automatique, un bot ne sera performant que si sa base de données est précise, organisée et améliorée continuellement. C'est pourquoi, avant toute chose, nous avons envisagé cette réorganisation technique. Dans ce cadre, je me suis appliquée à donner au métier des astuces et bonnes pratiques dans la gestion de la base de connaissances.

Dans un premier temps, je me suis interrogée sur l'organisation des connaissances du robot conversationnel. J'ai pu identifier des incohérences dans le rangement de certaines connaissances, notamment au niveau du classement en thématiques. J'ai également identifié plusieurs connaissances obsolètes, suite à des évolutions du contexte ou de nouvelles réglementations. Il est important de vérifier la pertinence et la conformité des informations. Certaines connaissances pouvaient être supprimées ou rassemblées en cas de redondance.

Dans un second temps, il a été essentiel de relire l'historique des conversations entre les utilisateurs et l'assistant virtuel. En effet, ce travail permet de suivre la satisfaction des usagers et l'évolution de leurs besoins. A cette étape, j'ai donc analysé les conversations du chatbot EDF Recrute, en apportant une attention particulière à celles dites en échec. On appelle conversation en échec, un dialogue terminé par une incompréhension du bot, c'est-à-dire lorsqu'il n'a pas pu satisfaire la requête de l'utilisateur. Dans ce cas, nous pouvons nous demander s'il manque des informations au sein de la base ou bien si la réponse est déjà présente mais qu'elle peut être complétée pour améliorer la compréhension du robot.

Par mes observations faites sur la base de connaissances et sur les dialogues, j'ai soumis plusieurs pistes de réorganisations au métier. Ce dernier a désormais les clés en main pour faciliter la compréhension et l'accès aux informations contenues dans le bot.

2.2.2.3. Travail réalisé sur le Design

En parallèle de la mise à jour de la base de connaissances du chatbot EDF Recrute, nous avons également préparé sa transformation Design. A travers le mot « Design », nous entendons la partie visible de l'application, celle qui sera manipulée par l'utilisateur. L'ensemble de ces évolutions jouent aussi un rôle important pour rendre le bot plus attractif.

Contexte

Comme précédemment expliqué (cf. 2.2.2.1 Contexte), nous avons plusieurs fois échangé avec le bot manager d'EDF Recrute sur les possibilités de modernisation, notamment suite à mon analyse comparative des chatbots dédiés au recrutement. De ce fait, nous avons peu à peu identifié des pistes d'amélioration qui pourraient être intéressantes.

Pistes d'amélioration
Identité du bot : avatar, couleurs, nom et personnalité (exemples : ton à adopter, emoji...)
Proximité utilisateur : tutoiement, stimulation visuelle (exemples : notification, envoi d'un message si l'utilisateur est inactif...)
Pro-activité : expliciter le rôle de l'assistant, meilleure prise en charge de l'utilisateur, dynamisme
Interface : retravailler la chatbox (fenêtre conversationnelle), aérer les dialogues, ajouter un carrousel
Fonctionnalités supplémentaires : CV matching, Mode vocal, matching offres d'emplois.

Figure 12: Pistes d'amélioration envisagées

Équipes mobilisées

MOE	MOA
Pôle Design d'Expérience (UX) du groupe Développement Agile de Solutions Lab : deux experts	Pôle Mobilité et Recrutement de la DRH Groupe : la Bot Manager EDF Recrute (chef de projet digital marque employeur) et sa manager
Pôle Solutions IA Éditeurs : un chef de projet - moi-même, supervisée par ma tutrice	

Figure 13: Composition de l'équipe projet

La composition de l'équipe du projet de modernisation EDF Recrute regroupe plusieurs acteurs. Le pôle UX (User eXperience) Design s'attache à concevoir et offrir une expérience utilisateur optimale, la meilleure possible. Le pôle Éditeurs travaille régulièrement avec le pôle Design afin de construire l'identité d'un bot et veiller à l'ergonomie de la solution.

Atelier de cadrage

Une fois les besoins métier identifiés et les solutions envisagées, il était important d'organiser un atelier design mobilisant les acteurs du projet afin de définir plus clairement

le cadre de l'étude et disposer d'un plan de réalisation. Pour ce faire, je me suis attelée à sa préparation en contactant le pôle UX Design pour lui présenter le contexte et son enjeu. J'y ai aussi présenté les études comparatives que j'avais menées auparavant. Par la suite, j'ai souvent échangé avec le pôle UX Design pour préparer l'atelier que nous avons co-animé.

Le jour de l'atelier design, l'ensemble des acteurs du projet a été sollicité pour donner une nouvelle image au chatbot, plus moderne. Ensemble, nous avons réfléchi à un nouvel avatar, une nouvelle personnalité, avons évoqué plusieurs idées de nom mais aussi travaillé sur une nouvelle charte graphique. De plus, nous nous attachons à respecter des normes d'accessibilité dans le but de rendre le chatbot utilisable par tous. Par ce biais, nous nous sommes par exemple entendues sur une identité de type « friendly », avons préféré le vouvoiement et nous sommes arrêtées sur le nom « AR-IA » pour Aide au Recrutement et Intelligence Artificielle. Aussi, nous avons mené une séance de brainstorming⁵ sur l'ensemble des fonctionnalités que nous proposons dans le cadre de l'offre de service SAVE, basée sur la solution de DYDU. De ce fait, nous nous sommes penchées sur diverses questions, notamment sur le paramétrage des fonctionnalités. Nous avons passé en revue celles préexistantes mais aussi évoqué de nouvelles fonctions dont le chatbot actuel ne bénéficie pas aujourd'hui. En effet, depuis la création du chatbot en 2018, plusieurs fonctionnalités ont été retravaillées par l'éditeur de logiciel et d'autres ont été créées. De nouvelles versions de l'interface utilisateur ont aussi été récemment proposées par l'éditeur et donc désormais disponibles pour cette refonte de chatbot.

Fonctionnalités	Explications
Teaser	Élément déclencheur qui permet d'afficher la boîte de dialogue pour interagir avec le chatbot (cf. Fig.15)
Disclaimer RGPD	Donne des informations et précisions sur la politique RGPD (Règlement Général sur la Protection des Données)
Onboarding	Type 1 : Succession d'un certain nombre de fenêtres qui donnent des informations complémentaires sur bot, notamment sur son utilisation Type 2 : Sous forme d'un onglet spécifique (cf. Fig.16)
Header	Élément qui donne un accès direct à certaines fonctionnalités: l'impression, l'envoi par mail et autres informations complémentaires sur le bot (cf. Fig.16)
Chatbox	Fenêtre conversationnelle, c'est-à-dire l'élément qui permet d'interagir avec le bot. Exemple de paramétrage : taille, comptage des caractères, phrase d'accueil, top 3 des connaissances les plus appelées, champ d'envoi...
Carrousel	Élément qui permet de présenter de l'information à l'utilisateur (accès direct) sous forme de défilement d'images et/ou d'informations (cf. Fig.16)

Figure 14: Brainstorming des fonctionnalités de l'atelier Design

Balayer l'ensemble de ces fonctionnalités nous a permis d'affiner les évolutions à réaliser dans le cadre du projet EDF Recrute. Cette étape est cruciale car elle nous permet de

⁵ Anglicisme désignant une technique de recherche d'idées originales dans une réunion, chacun émettant ses suggestions spontanément. Source : Le Robert.

préciser le projet tout en s'assurant que l'ensemble des équipes dispose de la même compréhension du besoin. Le livrable de cet atelier est un document présentant l'organisation du projet. J'ai eu l'occasion de réaliser ce compte-rendu et de veiller à sa conservation en tant qu'archive du projet. La prochaine étape du processus de cadrage du projet consiste à réaliser une maquette réunissant l'ensemble des éléments précédemment étudiés puis validés par toute l'équipe projet.

Réalisation

L'étape du maquettage est prise en charge par le pôle UX Design. Les experts travaillent alors à la fois sur le dessin du nouvel avatar du bot mais aussi sur l'agencement design de l'ensemble des fonctionnalités que nous avons validées lors de l'atelier. Une fois la maquette finalisée, les experts me l'ont transmise afin que je puisse la présenter au métier pour la validation finale.



Figure 15: Teaser EDF Recrute



Figure 16: Onboarding type2 EDF Recrute

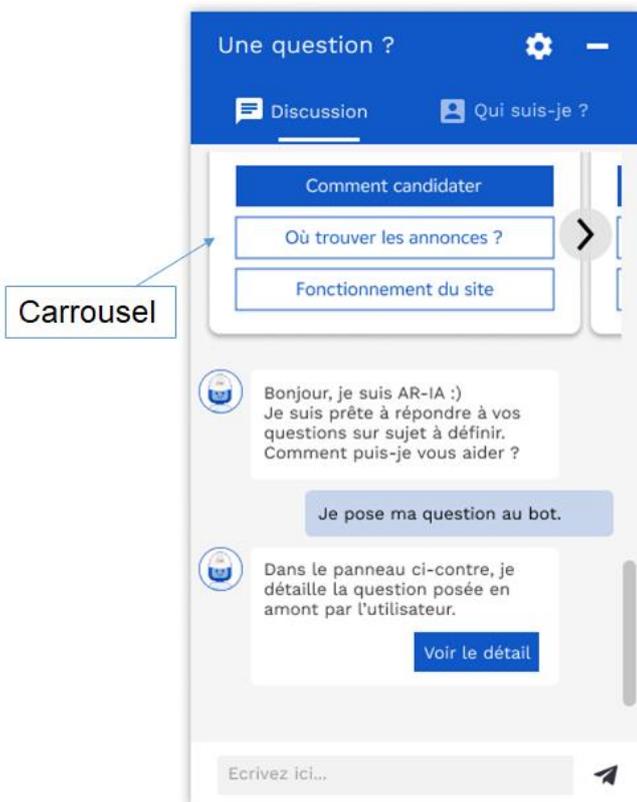


Figure 17: Chatbox EDF Recrute

Actuellement, la maquette est toujours en cours de validation car certains détails, notamment sur l'avatar du chatbot sont encore à retravailler. En effet, l'étape de l'avatar doit être validée par plusieurs départements de la communication EDF étant donné que le chatbot est ouvert à l'externe. Il est donc primordial que le logo soit en accord avec l'image que le Groupe EDF veut renvoyer. En parallèle, je continue d'accompagner le métier dans la refonte de la base de connaissances. Il m'a d'ailleurs actuellement sollicitée pour l'aider à travailler sur les formulations des connaissances et des réponses du bot.

Je viens de vous présenter plusieurs services assurés par le pôle Éditeurs dans le cadre du maintien en conditions opérationnelles d'une solution de l'offre de service SAVE. En particulier avec le chatbot EDF Recrute, je suis intervenue sur les points suivants : accompagnement du métier par un support technique et fonctionnel et amélioration continue du bot. A travers différentes missions effectuées sur différents chatbots, je suis intervenue sur d'autres activités qui relèvent de la MCO telles que les tests de non-régression lors de montées de versions de la solution ou les tests de la fonctionnalité vocale d'un voicebot.

2.3. Enrichissement de SAVE à travers un cas d'usage : le chatbot EDF Hydro

La partie précédente de ce rapport nous a permis d'examiner la mise en œuvre de l'offre de service SAVE en phase de « Run ». Toutefois, comme nous pouvons le voir sur la figure 7 (cf. p14), SAVE assure également les projets en phase de « Build », c'est-à-dire la mise en place de nouveaux chatbots sur des périodes de 3 à 6 mois. C'est pourquoi, nous allons désormais nous focaliser sur cette phase à travers le projet EDF Hydro. Nous verrons aussi qu'il peut s'avérer utile d'enrichir l'offre de service par le biais de développements internes.

2.3.1. Naissance du projet chatbot EDF Hydro

2.3.1.1. Contexte

L'énergie hydraulique est la première des énergies renouvelables en France. Comme nous l'avons précédemment présenté (cf. 1.1 Présentation de l'entreprise), EDF mise sur la montée en puissance des énergies renouvelables. Ainsi, la Direction de la Production et de l'Ingénierie Hydraulique, plus classiquement nommée EDF Hydro, est en charge du parc de production hydraulique d'EDF en France. Elle se compose notamment de six unités de production hydraulique réparties sur le territoire français.

EDF Hydro s'attache à servir la performance industrielle mais aussi environnementale. En effet, elle pratique l'économie circulaire de par sa volonté à limiter sa consommation et le gaspillage des ressources, en réduisant sa production de déchets mais aussi en les valorisant, les rendant utilisables par d'autres acteurs. Par conséquent, la gestion

des déchets est une tâche incontournable dont la réglementation se veut de plus en plus draconienne et dont les ambitions du Groupe ne cessent de croître.

Hydro Centre, l'une des six unités de production d'EDF Hydro, a sollicité la DSIT dans le but de mettre en place un assistant virtuel pour la prévention et la gestion des déchets au service de la simplification et pour tous ses salariés. Ce projet de chatbot est inédit et audacieux car ce type d'outil n'existe pas aujourd'hui pour les déchets industriels.

2.3.1.2. Validation des besoins

Fonctionnement actuel

La gestion des déchets industriels demeure au cœur des préoccupations d'EDF Hydro Centre. Toutefois, le métier nous a confié les contraintes fortes devant l'abondante documentation et la complexité de la réglementation déchets engageant la responsabilité pénale du directeur de sous-unité. EDF est reconnue pour sa rigueur dans le domaine des déchets, mais elle souffre d'un manque de simplicité de son processus opérationnel. En matière de gestion des déchets, la documentation est complète certes, mais fournie, très dense et plutôt compliquée. En effet, elle regroupe de nombreux guides de valorisation, prévention, réemploi, réglementaires etc. De plus, la réglementation évolue vers toujours plus d'exigences, notamment en matière de valorisation des déchets. De ce fait, la rigueur demandée en matière d'entreposage, de transport et de traçabilité, oblige les correspondants déchets à consulter sans cesse l'ensemble de ces bases de données, parfois hébergées à différents endroits (bases Sharepoint, sites intranets etc.).

Ainsi, l'unité de production, composée d'experts du domaine (ingénieurs, correspondants déchets etc.) doit faire face à de nombreuses sollicitations de la part des acteurs sur le terrain, au niveau des centrales hydrauliques. Ces derniers s'interrogent sur l'activité de gestion de leurs déchets industriels. Or, la gestion documentaire est devenue compliquée et chronophage. De plus, l'unité fait appel à une prestation externe pour répondre aux exigences en matière de gestion de déchets industriels, en particulier des déchets dangereux.

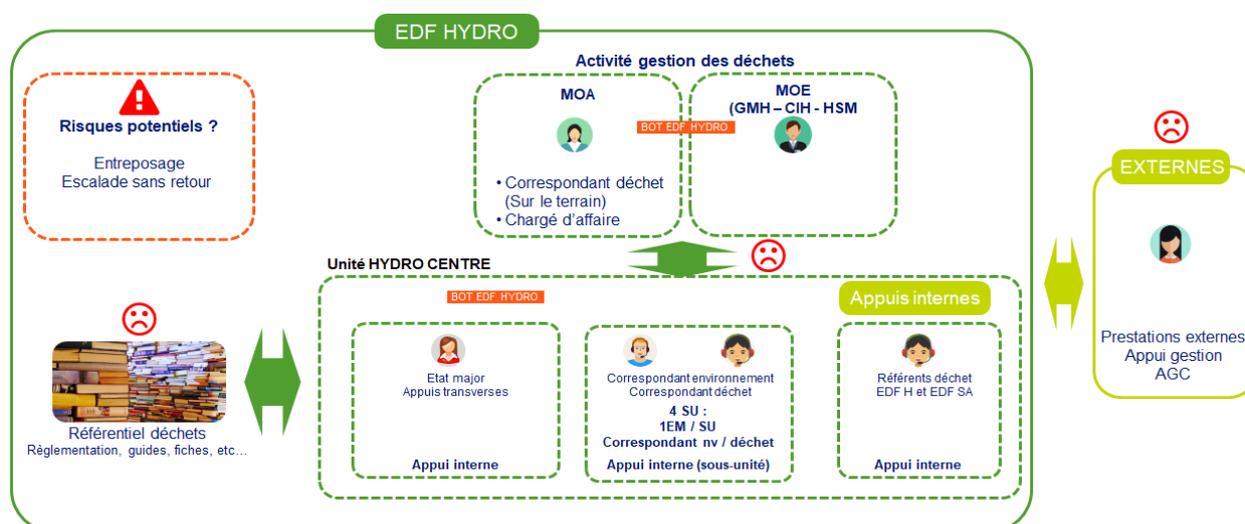


Figure 18: Contexte actuel de la gestion des déchets industriels Hydro Centre

Il est également important de préciser que les correspondants déchets des unités de production sont très peu formés à la réglementation déchets car ce rôle n'est qu'une « casquette » parmi d'autres, une mission annexe à leur métier de base.

Solution

Au vu de ces nombreuses sollicitations et de la gestion documentaire chronophage que représente l'activité de gestion des déchets industriels, la mise en place d'un assistant virtuel paraît judicieuse pour renforcer la **prévention**, la **gestion** et le **traitement** des déchets industriels d'EDF Hydro.

Bénéfices de l'utilisation d'un assistant virtuel :

- Sur le plan stratégique : informations centralisées au sein d'un seul outil. Tous les collaborateurs d'EDF Hydro, que ce soit les agents sur le terrain ou les correspondants de l'unité de production, peuvent plus facilement accéder aux informations. Le robot permet d'élargir le périmètre informatif : aspect réglementaire mais aussi préventif.
- Sur le plan opérationnel : gain temporel lié à un accès plus rapide à l'information et face à la prévention des erreurs. Le bot facilite l'activité des experts en réduisant les sollicitations directes.
- Sur le plan humain : innovation qui concerne un domaine peu valorisé/valorisant. Elle associe EDF mais aussi le domaine de la gestion des déchets à une image de modernité.
- Sur le plan financier : le métier entend maîtriser des coûts liés au recentrage des sollicitations de la prestation externe sur des sujets à forte plus-value mais aussi lié à des OPEX récurrentes, à des coûts évités, à la maximisation du rachat des matières valorisables et à un meilleur respect des exigences extra-financières du Groupe.

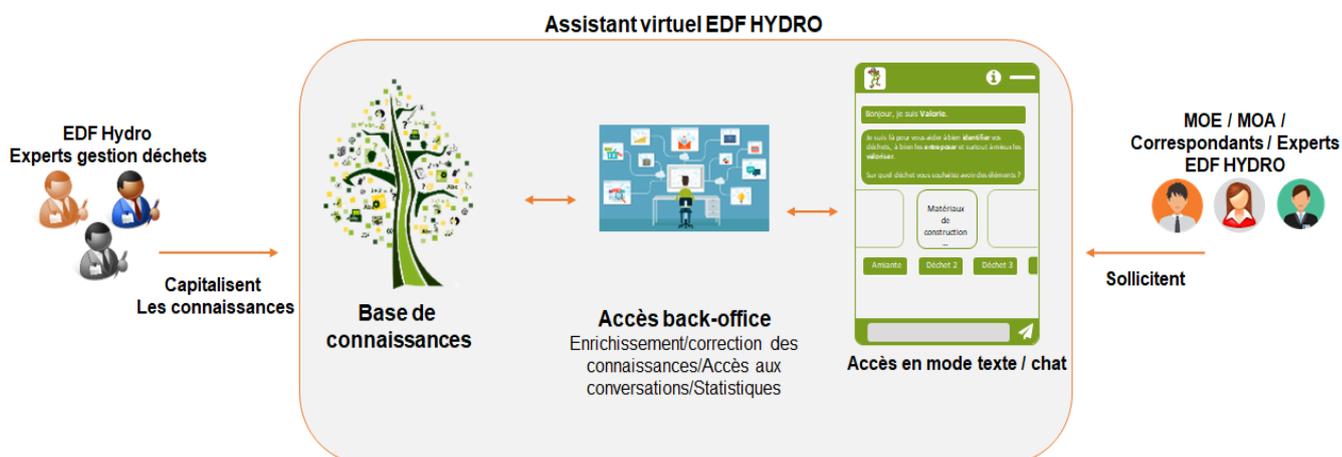


Figure 19: Validation du besoin

2.3.1.3. Organisation du projet

Équipes mobilisées

MOE	MOA
Pôle Solutions IA Éditeurs : chefs de projet, direction de projet IA, référent technique	EDF Hydro Centre : responsable produit, commanditaire, utilisateurs clés, experts métiers
Solution de l'éditeur DYDU : expert	
Pôle Design d'Expérience (UX) : experts	

Figure 20: Constitution de l'équipe projet EDF Hydro

Le responsable produit, aussi appelé Product Owner, a pour mission de maximiser la valeur d'affaire du chatbot. Il va alors véhiculer la vision produit du métier, identifier, prioriser et clarifier les fonctionnalités nécessaires. Il aura aussi pour tâche de formaliser les connaissances en concertant les experts métiers et les utilisateurs clés. Enfin il aura la décision finale sur le travail effectué. Les experts métiers sont les points de contact privilégiés du Product Owner car ils lui retranscrivent leurs visions métier.

Planning



Figure 21: Planning découpé en quatre phases

2.3.2. Déroulement de la phase de cadrage

La phase de cadrage du projet EDF Hydro s'est étendue sur quatre semaines. Elle a débuté par une réunion de lancement, aussi appelée Kick Off, qui permet de réunir l'ensemble des acteurs du projet et de s'assurer de la bonne compréhension du besoin par tous. Cette réunion permet aussi de présenter la démarche du projet au métier. Ensuite, ce lancement est suivi de trois ateliers de cadrage.



Figure 22: Découpage de la phase de cadrage

Atelier Design

Le premier atelier correspond à l'atelier Design (cf. 2.2.2.3). Lors de cette réunion, tous les acteurs du projet sont réunis afin de donner une identité au chatbot. Nous avons choisi le nom VDD'Bot pour Valorisation Des Déchets avec comme visuel une dame grenouille (avatar déjà existant chez EDF Hydro). Sa personnalité sera en adéquation avec la stratégie éditoriale et il adoptera le tutoiement. Il s'exprimera dans un langage courant avec un style directif. Le ton sera professionnel, cordial et bienveillant (telle une assistance technique). Les connaissances de type humour seront donc désactivées. Aussi, le bot préférera la vulgarisation au langage d'expert. Par ailleurs, l'entité graphique sera en adéquation avec la charte graphique du Groupe et dans le respect des normes d'accessibilité. Après validation de l'identité du bot, une maquette est proposée par le pôle UX Design.



Figure 23: Avatar

Ateliers Connaissances

Un objectif primordial de l'atelier connaissances est l'identification de toutes les sources de données nécessaires à la conception de la base de connaissances. Etant donnée la spécificité de la documentation dont fait l'objet la gestion des déchets chez EDF Hydro Centre, il nous a fallu anticiper cet atelier. C'est pourquoi, sur la base des documents transmis auparavant par le métier, mon collègue chef de projet et moi-même avons préalablement envisagé une stratégie de priorisation, de recueil et d'organisation des données.

La priorisation des données : un premier document de référence a été désigné comme source principale d'alimentation de la base. Ce guide rassemble un ensemble de fiches déchets propres à EDF et résume la réglementation, les dispositions applicables à chaque type de déchets ainsi que les voies de gestions possibles. D'autres sources de données ont également été étudiées pour la phase d'enrichissement des connaissances. En parallèle, les historiques des sollicitations d'utilisateurs pourront compléter l'enrichissement de la base. Cette approche a été présentée au métier qui l'a validée lors de l'atelier.

Un autre enjeu dans la priorisation des données demeure d'envisager des regroupements de connaissances par thématiques. Dans le cadre de ce projet spécifique, nous avons identifié trois thématiques en fonction d'un code présent sur les fiches déchets. Elles sont à prioriser pour la phase de réalisation car nous procéderons à une injection progressive par « lotissement » ventilés sur plusieurs sprints. Cette démarche relève de la méthode agile, elle permet d'impliquer le métier dès le départ mais aussi de lui fournir des éléments livrables à la fin de chaque sprint.

Le recueil des données : nos premières études des données nous ont convaincus que nous n'avons pas affaire à une phase de « Build » habituelle. La chaîne de traitement classique consiste à accompagner le métier dans la construction de la base de connaissances. Ce dernier rassemble l'ensemble des connaissances et les organise dans un fichier excel que

le pôle Éditeurs injecte dans l'outil. Enfin, les deux acteurs optimisent et enrichissent ensemble la base de connaissances (ajout de pièces jointes, liens éventuels etc.).

Toutefois, nous ne pouvons procéder de cette manière pour le projet EDF Hydro qui est un cas très spécifique. Tout d'abord, il s'agit d'un chatbot qui va capitaliser des connaissances pour diffuser de l'information précise, dense et plus ou moins complexe. Aussi, l'étude du document de référence nous a alertés sur la quantité très élevée d'informations à prélever, et ce pour chaque fiche déchet. Nous nous sommes donc interrogés sur l'investissement humain et temporel que ce travail d'extraction manuel demanderait. De plus, nous avons remarqué que la structure de chaque fiche déchet était répétitive. C'est pourquoi, après plusieurs concertations, nous avons opté pour une extraction automatique des données. Ce processus automatisé, développé en interne, est singulier et ne correspond pas à la chaîne de traitement usuelle. Il s'agit d'un service inédit proposé dans le cadre de l'offre de service SAVE.

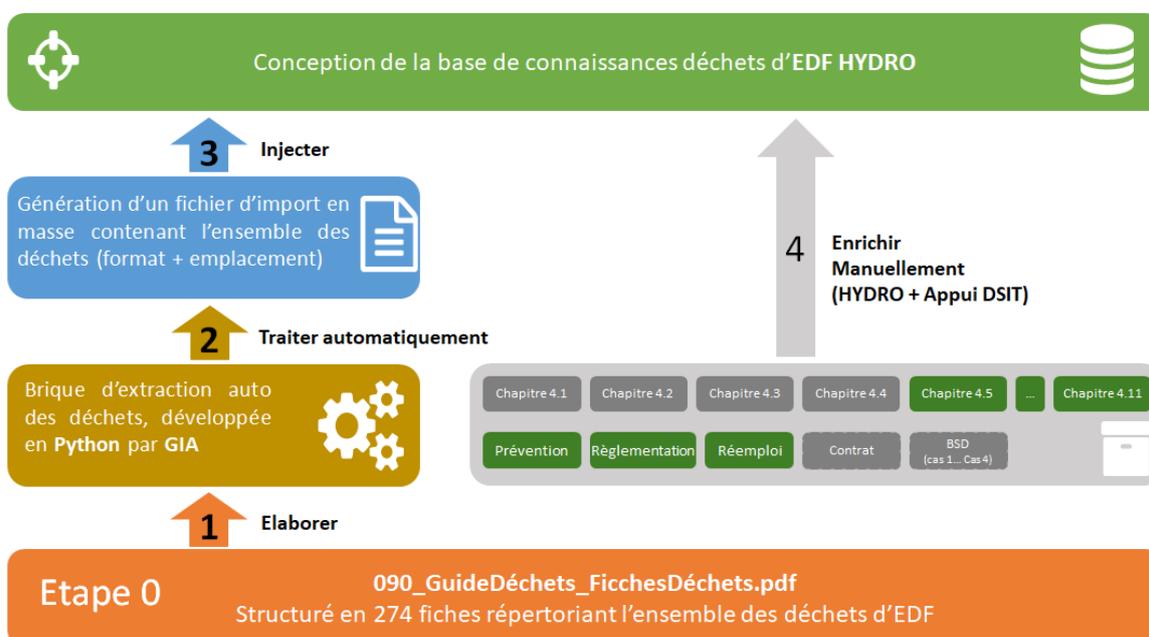


Figure 24: Chaîne de traitement mise en place pour EDF Hydro

L'organisation des données : la présentation des données est capitale et contribue fortement à une expérience utilisateur optimale. Lors d'une séance de brainstorming, nous avons envisagé de séparer les déchets en deux catégories : les déchets simples et les familles de déchets qui correspondent à des matériaux que l'on peut retrouver dans plusieurs déchets simples. La présentation de ces deux catégories de déchets sera réalisée différemment.

1. Un déchet de type simple sera présenté à l'utilisateur du bot sous la forme d'une carte d'identité. Nous y retrouvons l'ensemble des informations extraites à partir de sa fiche déchet. Cette identification sera présentée sous forme d'un panneau latéral associé à la chatbox, où l'ensemble des informations sera structuré (cf. 2.3.3 Automatisation). Ainsi, la réponse est apportée à l'utilisateur de manière immédiate et complète.

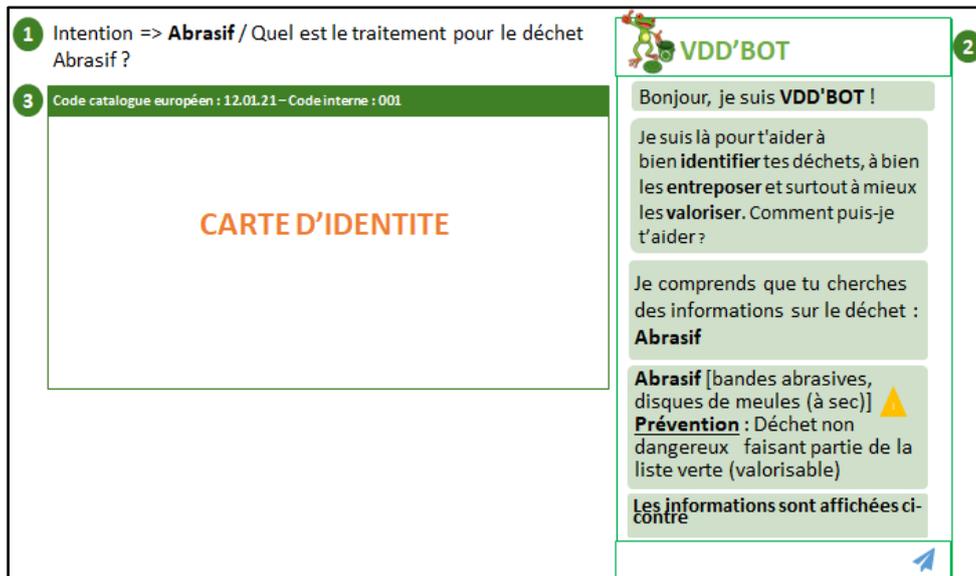


Figure 25: Organisation des données pour un déchet simple

2. Un déchet de type complexe, ou autrement dit famille de déchets, renverra à des liens de redirection vers d'autres connaissances, c'est-à-dire vers des déchets de type simple, potentiellement ciblés par l'utilisateur. Cette organisation permet d'affiner la requête de l'utilisateur, notamment s'il n'a pas identifié la nature précise du déchet dont il dispose ou sur lequel il veut se renseigner.



Figure 26: Organisation des données pour une famille de déchet

Atelier Intégration

Le dernier atelier de cette phase de cadrage permet d'envisager le contexte d'intégration de la solution. Plusieurs sources d'intégration ont été présentées et comparées, telles que la page web blanche classique et différentes applications d'entreprises. Nous avons convenu que l'expérimentation sera lancée en simultanée sur les 3 solutions suivantes : un

lien pointant vers le bot sur la communauté intranet EDF Hydro Centre, un espace dédié sur SharePoint et l'application OGIDE consacrée aux registres réglementaires. Le fait de déployer le bot sur une application déjà utilisée par les métiers permet de toucher davantage d'utilisateurs et ce de façon immédiate et sans communication dédiée nécessaire. En cas de succès de l'expérimentation, le périmètre de déploiement sera élargi.

Restitution

La phase de cadrage se termine par un bilan réalisé entre les équipes du projet. Elle permet de décider de la poursuite des travaux et de l'engagement de la phase suivante.

2.3.3. Automatisation de la constitution de la base de connaissances

Mon année d'alternance étant dédiée au TAL, le pôle Éditeurs m'a confié la mission d'extraction automatique des données qui vont peupler la base de connaissances de VDD'Bot.

Langage de programmation et librairies

Dans le cadre de cette extraction automatique, j'ai réalisé un développement en langage de programmation Python. J'ai pris l'habitude de travailler en local, en utilisant l'outil Jupyter Notebook de l'environnement de développement Jupyter. J'ai utilisé à des librairies libres de droit et traditionnelles dans les domaines de la data science et du traitement automatique des langues :

- **regex** : module standard Python qui permet de manipuler des expressions régulières.
- **pdftotext** : permet d'extraire du texte à partir d'un fichier PDF simple.
- **PyMuPDF** (alias **fitz**) : permet de visionner des fichiers aux formats PDF.
- **pandas** : permet la manipulation et l'analyse des données.
- **numpy** : permet de manipuler des matrices ou tableaux multidimensionnels.

Analyse du document

Le document identifié comme le socle de la base de connaissances du chatbot m'a été envoyé au format pdf. Étant donné qu'un fichier de ce format n'est pas toujours simple à manipuler, j'ai dû commencer par vérifier qu'il était modifiable. S'il ne l'est pas, les librairies python dédiées à ce format demeurent insuffisantes et ne peuvent extraire des informations de son contenu. Le cas échéant, il nous faut alors procéder à de l'océrisation, une technique de reconnaissance optique des caractères. Le guide référentiel des déchets étant au format pdf modifiable, cela a grandement facilité la suite des opérations.

Je me suis ensuite concentrée sur sa structure. Le document rassemble 274 fiches déchets à partir desquelles j'ai étudié la disposition des informations. Cette étape m'a permis de préparer l'extraction automatique. Dès lors, j'ai constaté un format tabulaire aux bordures

invisibles, probablement dû à un format initial de type *Word*. Cette disposition m'a posé problème pour l'extraction automatique car la librairie **pdftotext** lit les informations de manière linéaire.

FILIERES DE TRAITEMENT ET EXEMPLES DE PRESTATAIRES			
Code D/R	Définition	Filières de traitement (correcte, acceptable ou interdite)	Exemple de prestataires
Valorisation			
R5	Recyclage ou récupération d'autres matières inorganiques	Correcte	

Figure 27: Extrait d'une fiche déchet au format pdf

```

FILIERES DE TRAITEMENT ET EXEMPLES DE PRESTATAIRES

Filières de traitement
Exemple de
Code D/R Définition (correcte, acceptable
prestataires
ou interdite)

Valorisation

Recyclage ou récupération d'autres
R5 Correcte
matières inorganiques
    
```

Figure 28: Lecture par la machine de la fiche Fig.27

Les deux précédentes illustrations nous permettent de constater la différence entre les données présentées sur le document pdf et la manière dont la machine les lit et les imprime. Par exemple, la définition du traitement « Recyclage ou récupération d'autres matières inorganiques » est scindée en deux et intercalée avec le code de valorisation « R5 » et la filière de traitement adaptée « Correcte ». Cette lecture linéaire des données est un facteur important à considérer pour la suite de mon travail.

Structure des données

Cette étape m'a permis de mettre en lumière les informations que nous voulions extraire à partir des rubriques d'une fiche déchet. Pour cela, nous avons organisé un point avec le métier qui nous a précisé quelles données étaient importantes pour la constitution de la base de connaissances du chatbot. Effectivement, le métier étant au plus proche des utilisateurs cibles, il est le plus à même de savoir quelles informations sont pertinentes ou non. Néanmoins, la plupart des données présentes sur les fiches déchets nous permettra de constituer la carte d'identité d'un déchet.

Sur la figure 29 présentée ci-dessous, les informations entourées en vert correspondent à des données toujours présentes sur l'ensemble des 274 fiches déchets. La colonne « Exemple de prestataires » n'est pas à extraire car ces données sont pour la plupart obsolètes. Les rubriques « Observations » et « Précision », entourées en rouge, sont des

données facultatives et ne sont pas systématiquement présentes sur une fiche déchet. Toutefois, nous avons décidé de les extraire pour compléter la carte d'identité d'un déchet.

Nature du déchet	PEINTURE, ENCRE, COLLES ET RESINES NE CONTENANT PAS DES SUBSTANCES DANGEREUSES		
<i>Précision</i>	<i>Emballages vides ou contenant des résidus secs exclusivement, se reporter aux fiches 145a et 146a.</i>		
Code catalogue européen :	20.01.28	Catégorie : DnDnl	Code interne : 226
DISPOSITIONS REGLEMENTAIRES			
Conditions d'entreposage sur site en attente de traitement	Sur rétention, dans le respect des règles de protection contre l'incendie.		
Bordereau de suivi de déchet	Obligatoire : Non Recommandé : Oui		
Transport (ADR)	Non soumis		
FILIERES DE TRAITEMENT ET EXEMPLES DE PRESTATAIRES			
Code D/R	Définition	Filières de traitement (correcte, acceptable ou interdite)	Exemple de prestataires
Valorisation			
R1	Utilisation principale comme combustible ou autre moyen de produire de l'énergie	Correcte	
Elimination			
D10	Incinération à terre	Acceptable	
OBSERVATIONS – AXE DE PREVENTION			
Prévention : Utilisation de produits éco-labellisés.			

Figure 29: Informations à extraire à partir d'un exemple d'une fiche déchet

Développement

L'identification des données à extraire étant faite, j'ai pu créer un script de développement à partir de zéro. J'ai commencé par définir plusieurs fonctions pour chaque type d'information que je voulais relever, c'est-à-dire pour chaque information encadrée en figure 29. Dans chaque fonction, j'ai utilisé des expressions régulières. Ces dernières correspondent à une suite de caractères typographiques, appelée aussi motif, décrivant un ensemble de chaîne de caractères. Les mécanismes de base pour former de telles expressions sont basés sur des caractères spéciaux de substitution, de groupement et de quantification. Par ce biais, j'ai pu filtrer l'ensemble des données présentes dans le fichier et retenir uniquement celles utiles à la construction d'une fiche d'identité d'un déchet.

La fonction présentée ci-dessous permet d'extraire la nature d'un déchet :

```
def get_nature_dechet(p):
    """Extraction de la partie déchet
    :param p: page
    :return : déchet
    :rtype : str
    """
    def clean_regex(text):
        """Nettoyage de l'extraction
        :param text: texte
        :type : str
        :return : texte nettoyé
        :rtype : str
        """
        text = re.sub(r"(N)?ature du( déchet)?", r"", text)
        text = re.sub(r"\bdéchet\b", r"", text)
        text = re.sub(r"\s+", r" ", text)
        text = text.strip()
        return text

    match_dechet = ""
    search_entete = re.search(r"Référence.*Indice.*Page.*", p, re.DOTALL)
    search_precision = re.search(r"\n( )?Précision.*", p)
    search_catalogue = re.search(r"\n( )?Code catalogue.*", p)
    if search_entete:
        match_entete = search_entete.group(0)
        if search_catalogue and not search_precision:
            match_dechet = re.sub(r"(Référence.*Indice.*Page.*?\n)(.*)(\n( )?Code catalogue)(.*)", r"\2", match_entete,
                flags=re.DOTALL)
            match_dechet = clean_regex(match_dechet)
            return match_dechet
        # Si présence d'une rubrique précision
        elif search_precision:
            match_dechet = re.sub(r"(Référence.*Indice.*Page.*?\n)(.*)([^\n]*[A-Z]\r\n( )?déchet)?(\r\n( )?[A-Z]*.*\n)?(.*)(\n( )?Précision)(.*)", r"\2", match_entete, flags=re.DOTALL)
            match_dechet = clean_regex(match_dechet)
            return match_dechet
        else:
            return match_dechet
    else:
        return match_dechet
```

Figure 30: Fonction pour extraire la nature d'un déchet

Dans cet exemple, la fonction contient plusieurs expressions régulières. Une première me permet de récupérer tout le contenu disponible après l'entête d'une fiche déchet. Deux autres expressions me permettent d'extraire les contenus correspondant à la rubrique des codes du déchet et à la rubrique « Précision », si elle existe. Ces deux extractions mises en variables me permettent de créer deux cas de figure possibles : si la fiche déchet ne possède pas de rubrique « Précision » alors je récupère la chaîne de caractères située entre la fin de l'entête et le titre « Code Catalogue ». En revanche, s'il existe une rubrique « Précision », alors je récupère le contenu positionné entre la fin de l'entête et le début du titre « Précision ». J'ai dû proposer plusieurs motifs de fin d'extraction à cause de la linéarité des informations. Enfin, la fonction *clean_regex()* me permet de nettoyer le contenu extrait par ma regex.

```
Nature du PEINTURE, ENCREs, COLLES ET RESINES NE
déchet CONTENANT PAS DES SUBSTANCES DANGEREUSES
Emballages vides ou contenant des résidus secs exclusivement, se reporter aux
Précision
fiches 145a et 146a.
```

Figure 31: Lecture de la rubrique « Nature du déchet » par la machine (cf. Fig.29)

```
PEINTURE, ENCREs, COLLES ET RESINES NE CONTENANT PAS DES SUBSTANCES DANGEREUSES
```

Figure 32: Extraction de la rubrique « Nature du déchet » (cf. Fig.29)

Toutefois, l'ensemble des rubriques à extraire ne m'ont pas demandé le même travail et la même méthode. J'ai d'ailleurs dû souvent réajuster mes expressions régulières en fonction des nombreuses exceptions présentes sur les fiches déchets. En effet, la disposition des informations est souvent variable d'une fiche à l'autre, bien que la structure soit la même. Ce fut particulièrement le cas avec l'extraction des filières de traitement, pour lesquelles j'ai dû répertorier l'ensemble des cas de figure de présentation des informations.

Code D/R	Définition	Filières de traitement (correcte, acceptable ou interdite)
Valorisation		
R1	Utilisation principale comme combustible ou autre moyen de produire de l'énergie	Correcte
Élimination		
D10	Incinération à terre	Acceptable

Valorisation
Utilisation principale comme combustible ou R1 Correcte autre moyen de produire de l'énergie
Élimination
D10 Incinération à terre Acceptable

Figure 33: Lecture par la machine

Figure 34: Rubrique « Filières de traitement » d'une fiche pdf

Dans l'exemple présenté en figure 34, la première partie de la définition de la Valorisation « Utilisation principale comme combustible ou » ainsi que la seconde « autre moyen de produire de l'énergie » sont scindées en deux. J'ai donc dû les extraire séparément puis les concaténer. J'ai également dû considérer les cas où une troisième partie de la définition était insérée à la suite de la filière de traitement voire entre le code et la filière de traitement.

Néanmoins, cette méthode n'a pu être appliquée pour l'extraction de la section « Élimination ». En effet, cette rubrique possède trop de cas de figure pour pouvoir recourir aux expressions régulières. C'est pourquoi, j'ai fait appel à un collègue data scientist du groupe IA qui m'a donné quelques conseils et outils pour traiter ce genre de cas. Nous avons donc fait appel à la librairie *fitz* et sa fonction *getTextWords()* afin de récupérer les coordonnées de chaque mot. Ces dernières permettent de reconstituer plus simplement et plus proprement les chaînes à extraire.

Mise en forme de l'extraction

La fonction *Main* permet d'appeler l'ensemble des fonctions d'extraction sur toutes les fiches déchets du document de référence. Une mise en forme de cet ensemble de données a donc ensuite été nécessaire. Pour ce faire, j'ai créé un data frame que je peux à la fin exporter en fichier excel. En effet, pour injecter une base de connaissances dans le portail d'administration d'un chatbot, nous devons respecter un certain modèle : fichier au format excel avec un nombre, un ordre et une appellation de colonnes spécifiques.

Nous avons donc réfléchi à la manière dont nous allons répartir les informations dans ce fichier. Étant donné que nous avons précédemment validé avec le métier l'organisation des données sous forme de carte d'identité d'un déchet (cf. Fig.25) nous n'avons plus qu'à répartir les informations entre le Panneau Latéral et la Chatbox.

Thématique	Sous-thématique	Question	Formulations	Réponse	Titre panneau latéral	Panneau latéral
« Fiches déchets »	Catégorie	Nature du déchet «simplifiée»	Nature du déchet + À définir par le métier	Nature du déchet + Précisions + Prévention	Code interne + Code catalogue européen	Conditions d'entreposage + Transport + BSD + Valorisation + Elimination

Figure 35: Répartition des informations au sein des colonnes type du fichier d'import

Les informations figurant en vert correspondent aux données que j'extrait automatiquement en utilisant des expressions régulières. En rouge sont les informations que je rajoute dans le développement en dehors de l'extraction de données. En effet, l'ajout de modifications directement dans le script nous évite de nombreuses manipulations manuelles dans la base de connaissances. Ces automatisations permettent de gagner en temps, en énergie et en productivité.

Principaux ajouts effectués automatiquement dans le data frame :

- **Colonne « Thématique »** : ajout de la thématique « Fiches déchets » à chaque déchet. Cela permet de regrouper toutes les connaissances sous cette thématique.
- **Colonne « Question »** : simplification du nom du déchet extrait par suppression des parenthèses et leur contenu. Cette manipulation permet donc de ne pas alourdir les noms des connaissances et de faciliter l'accès aux informations.
- **Colonne « Formulations »** : ajout de différentes formulations que le métier a identifiées et qui déclenchent une connaissance.
- **Colonnes « Réponse » et « Panneau latéral »** : ajouts au niveau design et parcours utilisateurs : balises HTML (écriture en gras, puces, codes couleurs) ou encore des « Split Bubble », c'est-à-dire la division du texte en plusieurs bulles conversationnelles.

Thématique	Sous-thématique	Question	Formulations	Réponse	Titre panneau latéral	Panneau latéral
269	Fiches Déchets	DI	VERRE (non alimentaire) VERRE\nQuel est le traitement pour VERRE ? nQuelle est la filière de traitement pour mon déchet VERRE ?nFaut-il un BSD pour VERRE ?nComment traiter le déchet VERRE ? nQuel est le code déchet pour VERRE ?nQuel est le code ADR pour VERRE ? nQuel est le code R pour VERRE ?nQuel est le BSD pour VERRE ?	VERRE (non alimentaire)<hr class="split" />Précision : Ce déchet est réglementé par le décret n°2016-288 du 10/03/2016 relatif aux déchets de papier, de métal, de plastique, de verre et de bois (aussi nommé déchets « 5 flux »). Pour les modalités d'application du décret, voir la fiche de position réf : D455516010913	Code interne: 172 - Code Européen déchet: 17.02.02	Conditions d'entreposage sur site en attente de traitement : Containers spécifiques Bordereau de suivi de déchet : Obligatoire : Non Recommandé : Oui Transport (ADR) : Non soumis Les filières de traitement: Valorisation <li role="listitem">R5 – Recyclage ou récupération d'autres matières inorganiques<ul role="list"><li role="listitem">CorrecteÉlimination Élimination Interdite

Figure 36: Extrait du data frame final pour le déchet Verre (non alimentaire)



Figure 37: Visualisation du même déchet côté interface utilisateur

Ci-dessus, la visualisation finale disponible pour l'utilisateur. Nous notons que les informations sont présentées et organisées comme voulu et que les balises ont bien été interprétées.

Automatisation des groupes de formulations

Les premières importations de connaissances dans le BMS nous ont conduit à mener des tests. Ils nous ont notamment permis de constater que le matching des requêtes était faible, particulièrement à cause de la longueur des noms de déchets. Néanmoins, il existe plusieurs astuces pour optimiser les connaissances et par conséquent les taux de réponse. Mettre en place des groupes de formulations en est une. Ces derniers agissent comme des dictionnaires de synonymes et permettent au bot de comprendre plusieurs phrases à l'aide d'une seule formulation. Il existe deux types de groupes de formulations.

Le premier correspond aux groupes de formulations de nature linguistique. Nous relevons par exemple des groupes de termes véhiculant le même sens, des catégories grammaticales ou encore des formats de date. Ces groupes sont inclus dans le BMS et constituent la base sociale du bot.

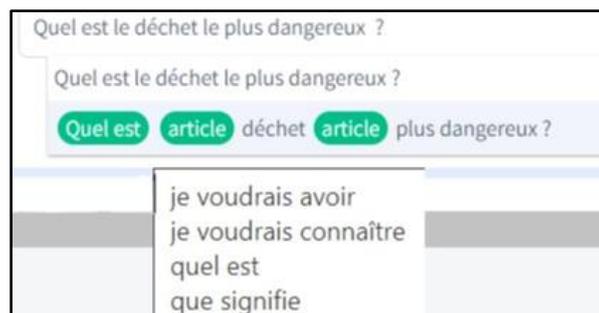


Figure 38: Exemples de groupes de formulation de type 1

Le second type de groupes de formulations correspond aux regroupements synonymiques que nous pouvons créer pour répondre à des besoins spécifiques. Dans notre cas, il nous a semblé pertinent de créer un groupe de formulations par déchet. Etant donné que les connaissances portent les noms simplifiés des déchets, nous jugeons utile de considérer les cas où un utilisateur questionnerait le bot à partir du nom complet du déchet. Pourtant, réaliser ces ajouts manuellement serait une tâche longue et fastidieuse. En effet, il est d'abord nécessaire de créer un groupe de formulations par déchet puis de l'appeler au sein de chaque formulation pour chaque connaissance. C'est pourquoi, une fois encore nous avons fait le choix d'automatiser ces deux étapes.

J'ai commencé par constituer automatiquement un fichier propre au format d'import de groupes de formulations dans le BMS. Il est constitué de deux colonnes : la première correspond au titre du groupe de formulations, à savoir un nom de déchet simplifié, tandis que la seconde contient les termes synonymes associés. Une fois le fichier importé, chaque groupe de formulation se voit attribuer un identifiant unique. Ce dernier peut être utilisé dans le code source de l'interface pour appeler le groupe de formulations associé. Par conséquent, j'ai dû récupérer l'ensemble des identifiants pour les inclure directement dans le script. Pour cela, il suffit de générer un fichier d'export des groupes de formulation. A partir de là, j'ai défini de nouvelles fonctions pour récupérer ces informations dans un data frame. Dès lors, j'ai pu remplacer dans le script original d'extraction automatique les noms des déchets par leurs identifiants. A cette étape, j'en ai aussi profité pour remplacer les motifs correspondant à des groupes de formulations linguistiques du BMS parmi les formulations d'une question, identifiées par le métier. Je les ai alors remplacés par leur code d'identification.

Dans l'illustration suivante, nous pouvons voir que tous les codes des groupes de formulations ont bien été ajoutés à la version originale présentée en figure 36.

Fiches déchets	DI	VERRE	<pre> \$_group_:66b86adc-2f66-4cf4-be34-c248bacc7912} \$_group_:d35bf94d-aecd-4f7b-80bf-728be707d9f1} le traitement pour \$_group_:66b86adc-2f66-4cf4-be34-c248bacc7912} ? \$_group_:d35bf94d-aecd-4f7b-80bf-728be707d9f1} la filière de traitement pour mon déchet \$_group_:66b86adc-2f66-4cf4-be34-c248bacc7912} ? Faut-il un BSD pour \$_group_:66b86adc-2f66-4cf4-be34-c248bacc7912} ? \$_group_:f563f377-3ea9-4384-bd25-59794e316c28} traiter \$_group_:7c978393-79c7-424b-b2fd-0df667c16f9d} déchet \$_group_:66b86adc-2f66-4cf4-be34-c248bacc7912} ? \$_group_:d35bf94d-aecd-4f7b-80bf-728be707d9f1} le code déchet pour \$_group_:66b86adc-2f66-4cf4-be34-c248bacc7912} ? \$_group_:d35bf94d-aecd-4f7b-80bf-728be707d9f1} le code ADR pour \$_group_:66b86adc-2f66-4cf4-be34-c248bacc7912} ? \$_group_:d35bf94d-aecd-4f7b-80bf-728be707d9f1} le code R pour \$_group_:66b86adc-2f66-4cf4-be34-c248bacc7912} ? \$_group_:d35bf94d-aecd-4f7b-80bf-728be707d9f1} le BSD pour \$_group_:66b86adc-2f66-4cf4-be34-c248bacc7912} ? </pre>	<pre> VERRE (non alimentaire)<hr class="split" />Précision : Ce déchet est réglementé par le décret n°2016-288 du 10/03/2016 relatif aux déchets de papier, de métal, de plastique, de verre et de bois (aussi nommé déchets « 5 flux »). Pour les modalités d'application du décret, voir la fiche de position réf : D455516010913 </pre>	<pre> Conditions d'entreposage sur site en attente de traitement : Containers spécifiques.

Bordereau de suivi de déchet : Obligatoire : Non Recommandé : Oui

 Transport (ADR) : Non soumis

Les filières de traitement

Valorisation
<li role="listitem">R5 – Interne: Recyclage ou récupération d'autres matières 172 - inorganiques<ul role="list"><li role="listitem">Correcte catalogu e: #FE5716;">Élimination
Élimination Interdite </pre>
----------------	----	-------	---	--	--

Figure 39: Extrait du data frame final avec ajout des groupes de formulations

Suite à ces ajouts d'informations automatisés, nous avons importé la nouvelle base de connaissances dans le BMS. Comme nous pouvons le constater sur la figure suivante, les codes sources des groupes de formulations ont bien été intégrés dans la plateforme. Les groupes apparaissent en couleur verte :



Figure 40: Affichage des groupes de formulations dans le BMS

2.3.4. Poursuite du travail

Le projet EDF Hydro est encore actuellement en phase de réalisation / de « Build ». Les prochaines étapes consistent à enrichir manuellement la base de connaissances à partir d'autres documents en matière de gestion, de traitement et de valorisation des déchets. Cet enrichissement est pris en charge par le métier avec accompagnement par le pôle Éditeurs. Aussi, de nouvelles extractions automatiques sont envisagées pour faciliter l'ajout des données d'enrichissement, parfois rébarbatif. J'envisage ainsi de poursuivre les développements internes en matière d'extractions d'informations. De plus, le projet entrera bientôt en phase d'expérimentation où un panel d'utilisateurs testera le chatbot et nous fera part de ses premières remarques. De nouveaux ajustements seront donc à prévoir. Enfin, si le projet est validé par le métier en fin de phase d'expérimentation, il sera mis en production auprès de toute la communauté Hydro Centre. Le pôle Éditeurs pourra alors proposer les différents services dispensés dans le cadre de la MCO, tels que l'accompagnement du métier et l'amélioration continue du chatbot.

Conclusion

Bilan

Nous nous étions demandé dans quelle mesure nous pouvions faciliter le travail d'un bot manager. Pour cela, nous avons exposé le cas d'usage d'EDF Recrute puisque ce projet répondait majoritairement à cet enjeu. En effet, de par la formation à la solution chatbot dispensée au métier, le support technique et fonctionnel, notamment dans la gestion de la base de connaissances mais aussi par l'amélioration continue du bot au cours de ce projet, je pense avoir montré plusieurs manières d'apporter un soutien au métier dans son rôle de bot manager.

Nous nous demandions également dans quelle mesure nous pouvions enrichir l'offre de service d'assistants virtuels à travers des techniques de TAL. En plus de participer au déploiement de l'offre de service depuis la naissance d'une idée jusqu'à la réalisation d'un chatbot, à travers le projet EDF Hydro, j'ai pu développer en interne un module spécifique. Ce dernier a permis de constituer automatiquement une base de connaissances puis de l'optimiser au maximum.

De manière plus générale, les différents projets auxquels j'ai participé m'ont permis de développer des compétences en matière de gestion de projet. De plus, j'ai eu la chance de participer à différents projets et d'intervenir sur des phases de « run » mais aussi des phases de « build ». Mes missions m'ont également beaucoup appris sur la relation client. En effet, je me suis entretenue avec différents métiers de différentes directions EDF auxquels j'ai dû m'adapter. J'ai compris l'importance du lien à entretenir avec le métier qui nous transmet un savoir à modéliser et pour lequel nous faisons en sorte de répondre à ses besoins. En outre, évoluer dans le groupe Intelligence Artificielle m'a permis de m'impliquer davantage dans l'univers stimulant de la data science.

D'un point de vue scolaire, j'ai pu au cours de ces derniers mois faire le lien avec différents cours que nous avons eus (gestion de projet, ingénierie des connaissances, programmation algorithmique et objet, langages réguliers etc.). Pouvoir les appliquer concrètement m'a donné l'occasion d'en saisir les enjeux.

Enfin, cette année d'alternance m'a appris autant sur le plan humain que professionnel. Évoluer dans une équipe pluridisciplinaire et dans un contexte sanitaire forçant le télétravail exige d'être à l'écoute des autres et de remettre en question ses approches. Je repars non seulement avec une vision riche de ce que peut offrir l'IA mais aussi avec la capacité de projeter ses applications concrètes dans le monde industriel.

Perspectives

Le développement interne que j'ai eu l'occasion de réaliser pour le chatbot EDF Hydro n'est pas le premier micro-service permettant d'enrichir l'offre de service SAVE. D'autres travaux ont déjà été menés auparavant avec pour objectifs d'améliorer et d'optimiser les services proposés par la plateforme mutualisée d'assistance virtuelle. Nous pouvons par exemple citer l'analyse des conversations des chatbots, l'implémentation de nouvelles statistiques à partir de la base de connaissances ou encore la correction d'anomalies détectées au sein de la base de connaissances. De tels outils engendrent des assistants virtuels plus robustes et plus performants. C'est pourquoi, nous envisageons de regrouper l'ensemble de ces micro-services au sein d'une même interface pour centraliser et faciliter l'exploitation de ces derniers par notre équipe. Ce travail s'inscrit dans la continuité des deux ambitions posées au début de ce rapport, à savoir faciliter le travail des bots managers et enrichir l'offre de service d'assistants virtuels déjà mise en place. La suite de mon alternance pourra notamment être dédiée d'une part à l'optimisation d'un de ces micro-services : l'analyse des conversations en échec et d'autre part à son intégration sur le portail mutualisé.

Annexes

Le manifeste de la raison d'être EDF est inclus dans l'affiche de communication ci-dessous :

**PERMETTRE QUE
LA VIE REDÉMARRE
MAIS PAS LE CO₂,
C'EST NOTRE
RAISON D'ÊTRE.**

Notre raison d'être est de construire un avenir énergétique neutre en CO₂ conciliant préservation de la planète, bien-être et développement, grâce à l'électricité et à des solutions et services innovants.

Indispensable à nos vies, l'énergie permet de s'éclairer et de se chauffer, de se nourrir, de se déplacer, de produire, de communiquer, d'apprendre, d'innover. De toutes les énergies, l'électricité, parce qu'elle est un bien de première nécessité, doit être accessible à tous et sur tous les territoires. C'est aussi l'énergie du progrès dans un monde de plus en plus numérique. Face à l'urgence climatique, nous voulons inventer, partout où nous sommes présents, un nouveau modèle énergétique : moins émetteur de CO₂, plus efficace, plus respectueux de l'environnement et des populations. Forte des valeurs de service public, EDF s'engage pour cet avenir énergétique juste, innovant et durable. En donnant à tous nos clients – particuliers, entreprises et territoires – la possibilité d'agir grâce à des services et des solutions pour moins et mieux consommer l'énergie, selon leurs besoins et leurs aspirations. En déployant au meilleur coût une électricité toujours moins carbonée, grâce au nucléaire et aux énergies renouvelables. En s'appuyant sur notre savoir-faire industriel et en recherchant sans cesse l'amélioration de nos performances. En nouant des partenariats avec l'ensemble des institutions, acteurs économiques et écosystèmes de l'innovation, en France et dans le monde. En encourageant le talent et l'engagement des femmes et des hommes qui font la force du groupe EDF. Car cette raison d'être, c'est notre raison d'avancer ensemble.

Devenons l'énergie qui change tout.



1552 001 317