Essai d'utilisation de l'intelligence artificielle pour rédiger des comptes rendus en médecine vétérinaire

Sandra Fourel

ANNEE 2024 THESE : 2024 – TOU 3 – 4110

ESSAI D'UTILISATION DE L'INTELLIGENCE ARTIFICIELLE POUR RÉDIGER DES COMPTES RENDUS EN MÉDECINE VÉTÉRINAIRE

THESE D'EXERCICE

pour obtenir le titre de
DOCTEUR VETERINAIRE

DIPLOME D'ETAT

présentée et soutenue publiquement
devant l'Université Paul-Sabatier de Toulouse
par

FOUREL Sandra, Emeline, Laura

Directrice de thèse : Mme Nathalie PRIYMENKO

JURY

PRESIDENT :
M. Jean-Philippe JAEG Maître de Conférences à l'Ecole Nationale Vétérinaire de TOULOUSE

ASSESSEURS :
Mme Nathalie PRIYMENKO Maître de Conférences à l'Ecole Nationale Vétérinaire de TOULOUSE
Mme Petra ROUCH-BUCK Ingénieure de recherche à l'Ecole Nationale Vétérinaire de TOULOUSE

MEMBRE INVITE :
M. Jean-Sébastien DESMEDT Chef de projet données de la recherche et intelligence artificielle

Ministère de l'Agriculture et de l'Alimentation
ECOLE NATIONALE VÉTÉRINAIRE DE TOULOUSE

Liste des directeurs/assesseurs de thèse de doctorat vétérinaire

Directeur : Professeur Pierre SANS

PROFESSEURS CLASSE EXCEPTIONNELLE

M. **BAILLY Jean-Denis**, *Hygiène et industrie des aliments*
M. **BERTAGNOLI Stéphane**, *Pathologie infectieuse*
Mme **BOURGES-ABELLA Nathalie**, *Histologie, anatomie pathologique*
M. **BOUSQUET-MELOU Alain**, *Pharmacologie, thérapeutique*
M. **BRUGERE Hubert**, *Hygiène et industrie des aliments d'origine animale*
M. **CONCORDET Didier**, *Mathématiques, statistiques, modélisation*
M. **DELVERDIER Maxence,** *Anatomie pathologique*
Mme **GAYRARD-TROY Véronique**, *Physiologie de la reproduction, endocrinologie*
M. **GUERIN Jean-Luc**, *Aviculture et pathologie aviaire*
Mme **HAGEN-PICARD Nicole**, *Pathologie de la reproduction*
M. **JACQUIET Philippe**, *Parasitologie et maladies parasitaires*
M. **MEYER Gilles,** *Pathologie des ruminants*
M. **SCHELCHER François,** *Pathologie médicale du bétail et des animaux de basse-cour*
Mme **TRUMEL Catherine**, *Biologie médicale animale et comparée*

PROFESSEURS 1ère CLASSE

Mme **CADIERGUES Marie-Christine**, *Dermatologie vétérinaire*
Mme **DIQUELOU Armelle**, *Pathologie médicale des équidés et des carnivores*
M. **DUCOS Alain**, *Zootechnie*
M. **FOUCRAS Gilles**, *Pathologie des ruminants*
M. **GUERRE Philippe**, *Pharmacie et toxicologie*
Mme **LACROUX Caroline**, *Anatomie pathologique, animaux d'élevage*
Mme **LETRON-RAYMOND Isabelle**, *Anatomie pathologique*
M. **LEFEBVRE Hervé**, *Physiologie et thérapeutique*
M. **MAILLARD Renaud,** *Pathologie des ruminants*
Mme **MEYNADIER Annabelle,** *Alimentation animale*
M. **MOGICATO Giovanni**, *Anatomie, imagerie médicale*

Mise à jour le 22/10/2024

PROFESSEURS 2ème CLASSE

Mme **BOULLIER Séverine**, *Immunologie générale et médicale*
Mme **CAMUS Christelle**, *Biologie cellulaire et moléculaire*
M. **CORBIERE Fabien**, *Pathologie des ruminants*
Mme **FERRAN Aude**, *Physiologie-Thérapeutique*
M. **MATHON Didier**, *Pathologie chirurgicale*
M. **NOUVEL Laurent,** *Pathologie de la reproduction*
Mme **PAUL Mathilde**, *Epidémiologie, gestion de la santé des élevages avicoles*
M. **VOLMER Romain**, *Microbiologie et infectiologie*

MAITRES DE CONFERENCES HORS CLASSE

M. **BERGONIER Dominique**, *Pathologie de la reproduction*
Mme **BIBBAL Delphine**, *Hygiène et industrie des denrées alimentaires d'origine animale*
M. **JAEG Jean-Philippe**, *Pharmacie et toxicologie*
Mme **LALLEMAND Elodie**, *Chirurgie des équidés*
M. **LIENARD Emmanuel**, *Parasitologie et maladies parasitaires*
M. **LYAZRHI Faouzi**, *Statistiques biologiques et mathématiques*
Mme **PALIERNE Sophie**, *Chirurgie des animaux de compagnie*
Mme **PRIYMENKO Nathalie**, *Alimentation*

MAITRES DE CONFERENCES CLASSE NORMALE

M. **ASIMUS Erik**, *Pathologie chirurgicale*
Mme **BRET Lydie**, *Physique et chimie biologiques et médicales*
Mme **BOUHSIRA Emilie**, *Parasitologie, maladies parasitaires*
M. **CARTIAUX Benjamin**, *Anatomie, imagerie médicale*
M. **COMBARROS Daniel,** *Dermatologie vétérinaire*
M. **CONCHOU Fabrice**, *Imagerie médicale*
Mme **DANIELS Hélène**, *Immunologie, bactériologie, pathologie infectieuse*
Mme **DAVID Laure**, *Hygiène et industrie des aliments*
M. **DIDIMO IMAZAKI Pedro**, *Hygiène et industrie des aliments*
M. **DOUET Jean-Yves**, *Ophtalmologie vétérinaire et comparée*
M. **FERCHIOU Ahmed,** *Economie et gestion des entreprises vétérinaires agricoles*
M. **FUSADE-BOYER Maxime,** *Microbiologie et infectiologie*
M. **GAIDE Nicolas,** *Histologie, anatomie pathologique*
Mme **GRANAT Fanny**, *Biologie médicale animale*
Mme **JOURDAN Géraldine**, *Anesthésie, analgésie*
M. **JOUSSERAND Nicolas**, *Médecine interne des animaux de compagnie*
Mme **LAVOUE Rachel**, *Médecine Interne*
M. **LE GRAVERAND Quentin,** *Alimentation animale*
M. **LE LOC'H Guillaume**, *Médecine zoologique et santé de la faune sauvage*
Mme **MEYNAUD-COLLARD Patricia**, *Pathologie chirurgicale*
Mme **MILA Hanna**, *Elevage des carnivores domestiques*
M. **OTAVIANO DO REGO Renato,** *Chirurgie*
Mme **PIERRON Alix**, *Pharmacie-Toxicologie*
M. **VERGNE Timothée**, *Santé publique vétérinaire, maladies animales réglementées*
Mme **WARET-SZKUTA Agnès**, *Production et pathologie porcine*

Mise à jour le 22/10/2024

INGÉNIEURS DE RECHERCHE

M. **AUMANN Marcel**, *Urgences, soins intensifs*
M. **AUVRAY Frédéric**, *Santé digestive, pathogénie et commensalisme des entérobactéries*
M. **CASSARD Hervé**, *Pathologie des ruminants*
M. **CROVILLE Guillaume**, *Virologie et génomique cliniques*
Mme **DIDIER Caroline**, *Anesthésie, analgésie*
M. **DELPONT Mattias**, *Clinique Aviaire*
Mme **DUPOUY GUIRAUTE Véronique**, *Innovations thérapeutiques et résistances*
Mme **GAILLARD Elodie**, *Urgences, soins intensifs*
Mme **GEFFRE Anne**, *Biologie médicale animale et comparée*
Mme **GRISEZ Christelle**, *Parasitologie et maladies parasitaires*
Mme **JEUNESSE Elisabeth**, *Bonnes pratiques de laboratoire*
Mme **LAYSSOL-LAMOUR Catherine**, *Imagerie Médicale*
Mme **POUJADE Agnès**, *Anatomie pathologique Vétérinaire*
Mme **PRESSANTI Charline**, *Dermatologie vétérinaire*
M. **RAMON PORTUGAL Felipe**, *Innovations thérapeutiques et résistances*
M. **REYNOLDS Brice**, *Médecine interne des animaux de compagnie*
Mme **ROUCH BUCK Pétra**, *Médecine préventive*
Mme **SAADA Chloé**, *Gestion intégrée de la santé des ruminant*

Mise à jour le 22/10/2024

REMERCIEMENTS

Au président du jury de thèse,

A monsieur le Docteur Jean-Philippe JAEG,
Maître de Conférences à l'école Nationale Vétérinaire de Toulouse,
Pharmacie et Toxicologie

Qui nous a fait l'honneur d'accepter la présidence de notre jury de thèse,
Pour l'intérêt porté à mon travail,
Sincères remerciements et hommages respectueux.

Au jury de thèse,

A Madame la Docteure Nathalie PRIYMENKO,
Maître de Conférences à l'Ecole Nationale Vétérinaire de Toulouse,
Alimentation et botanique appliquée.

Qui m'a accompagnée et conseillée tout au long de ce travail,
Pour votre réactivité et votre implication,
Sincères remerciements.

A Madame la Docteure Petra ROUCH BUCK,
Responsable des services cliniques de Médecine préventive et nutrition

Qui a très aimablement accepté de faire partie de notre jury de thèse
Pour les échanges stimulants et les conseils avisés lors de mes enseignements cliniques
Profonde gratitude.

A Monsieur Jean-Sébastien DESMEDT
Chef de projet données de la recherche et intelligence artificielle

Qui m'a conseillé et aidé dans la réalisation de cet écrit,
Profonde gratitude.

TABLE DES MATIÈRES

LISTE DE FIGURES

LISTE DES TABLEAUX

LISTE DES ANNEXES

LISTE DES ABREVIATIONS

IA : Intelligence artificielle

***ASR** : "*Automatic Speech Recognition"* - système de reconnaissance vocale automatique

***NLP** : "*Natural Language Processing"* - Modèle de traitement du langage naturel

***LLM** : "*Large Language Model"* - Grand Modèle du Langage

***RAG** : *"Retriever Augmented generation"* - Génération Augmentée par Récupération

APE : Antiparasitaire externe

API : Antiparasitaire interne

* Par souci de clarté, lors de la première apparition dans le texte, nous présenterons l'acronyme en français et en anglais puis l'abréviation anglaise sera la seule conservée car consacrée par usage.

INTRODUCTION

Le vétérinaire exerce de nombreuses fonctions : généraliste, imageur, chirurgien et doit effectuer des tâches administratives, peu passionnantes pour la plupart des professionnels. Parmi celles-ci figure la rédaction de comptes-rendus, une tâche récurrente qui a plusieurs intérêts : suivre l'animal en consignant les informations dans son dossier et disposer ainsi d'un historique détaillé, noter les informations à transmettre au propriétaire pour qu'il prenne soin au mieux de son animal après la consultation, coordonner les soins avec d'autres professionnels de santé lors de cas référés ou tout simplement entre vétérinaires d'une même structure durant des périodes de vacances et enfin, permettre la facturation. Ces écrits permettent aussi de prouver les actes réalisés et proposés ainsi que les traitements prescrits. Or, une étude récente a montré que les professionnels de santé passent environ 16 minutes par consultation sur les dossiers électroniques de santé de leurs patients, pour chaque rendez-vous. De même, les médecins passeraient environ 11% de leur temps sur les dossiers médicaux, en dehors de leurs heures de travail (Finnegan, 2020). Ces informations montrent le temps non négligeable alloué à la rédaction de ces derniers. De plus, l'arrivée de la génération Y sur le marché du travail s'accompagne d'une volonté d'intégrer l'utilisation des nouvelles technologies dans le quotidien vétérinaire, au profit de la qualité et de la rapidité des échanges (Février, 2017). Le développement d'outils digitaux s'inscrit donc dans une démarche de création d'un "vétérinaire augmenté", répondant aux souhaits des nouvelles générations et facilitant un retour au cœur du métier. Dernièrement, le développement de l'intelligence artificielle a permis une ouverture du champ des possibles à ce propos. Par exemple, une étude réalisée par l'université d'Harvard en Septembre 2023 dans une entreprise de conseil en gestion a mis en évidence une augmentation de la productivité et de la qualité du travail des consultants lorsque ces derniers intégraient l'intelligence artificielle dans la réalisation de certaines tâches. Concrètement, les individus sous le seuil de performance moyen augmentaient la leur de 43% et ceux au-dessus du seuil de 17%, respectivement (Dell'Acqua *et al.*, 2023). Ainsi, l'étude qui suit a pour objectif de décrire certains grands mécanismes de l'intelligence artificielle et la diversité des domaines où elle intervient actuellement en médecine vétérinaire. Il s'agit d'explorer en quoi la digitalisation du métier peut rendre plus confortable la part administrative du quotidien vétérinaire, pour ensuite

développer sur le projet de mise en place d'un outil permettant de rédiger facilement les comptes-rendus : ReqVet. En effet, par l'utilisation de certains outils d'intelligence artificielle, il est possible d'automatiser la rédaction de compte-rendu à partir de l'enregistrement vocal de la consultation et ainsi, générer rapidement un écrit avec une faible marge d'erreur ou d'omission.

Partie 1 - Etat des lieux : intelligence artificielle et médecine vétérinaire

L'intelligence artificielle intervient désormais dans de nombreux domaines et corps de métier. Concernant l'abord médical du monde du travail, elle a commencé par conquérir le marché de la médecine humaine. En effet, c'est dans les années 1960-1970 que l'IA fait son apparition dans le domaine de la santé avec un système d'aide au diagnostic (Lamoly, 2020). Dans le domaine vétérinaire, les premières applications de l'IA concernent les élevages avec des mangeoires intelligentes et le principe des "vaches connectées". Les technologies d'intelligence artificielle se sont ensuite étendues au secteur des animaux de compagnie avec pour cibles les vétérinaires en proposant des outils de détection précoce de certaines pathologies, mais également les propriétaires, en leur permettant de suivre de façon journalière l'évolution de leur animal sur de multiples points tels que le poids ou le temps de sommeil dans une journée et, ainsi, de se sentir plus proche d'eux.

L'objectif de cette thèse ne consiste pas à décrire en détail les outils pré-existants, déjà mis en valeur par certain.e.s de mes confrères et consoeurs (Lamoly, 2020; Amadieu, 2023; Mottay, 2023), mais simplement, dans un premier temps, d'en faire une liste actualisée afin de prendre pleine mesure de la place que l'IA occupe au sein de notre corps de métier.

1. Les domaines d'intervention de l'intelligence artificielle dans le monde vétérinaire

1.1. Aide à la décision avec ZAG®

ZAG® est un outil d'aide au diagnostic, au même titre que la recherche bibliographique, au bénéfice d'être plus vaste et plus rapide. C'est avec ce type d'outil que l'IA est d'abord apparue dans le monde vétérinaire permettant ainsi d'explorer toutes les hypothèses diagnostiques notamment dans le cas d'une activité de vétérinaire généraliste. En France, l'outil ZAG, créé en 2016 par l'équipe Pronozia puis racheté par Domes Pharma (anciennement TVM), répond à ce défi en réunissant toutes les informations sur l'animal : ses antécédents, l'anamnèse

rapportée par le propriétaire, les résultats d'examens complémentaires, pour les comparer ensuite à sa base de données et à la littérature. ZAG peut aussi compléter le diagnostic du vétérinaire avec des données pronostiques ou des pistes thérapeutiques et fournir des éléments bibliographiques pour corroborer les différentes hypothèses, permettant ainsi au vétérinaire de vérifier la fiabilité des sources et donc, des propositions de ZAG.

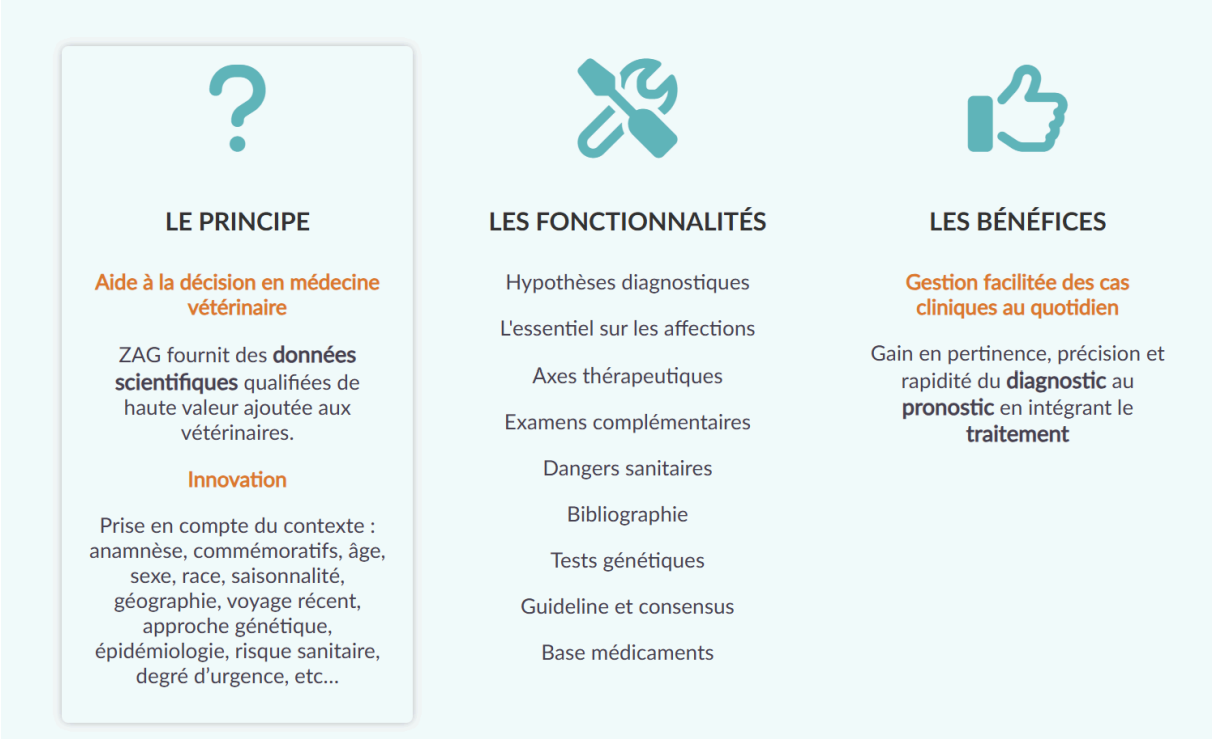


Figure 1 : Page d'accueil du site ZAG - zagbydomespharma.com (consulté le 27/05/2024)

Le service proposé par ZAG est simple : "gagner en pertinence, précision et rapidité" à tous les niveaux du diagnostic au pronostic en passant par le traitement. L'application regroupe plusieurs fonctionnalités afin de répondre à la notion de "vétérinaire augmenté" (figure 1). En effet, selon certaines études en santé humaine, l'exactitude diagnostic d'un professionnel de santé, seul dans le parcours décisionnel, atteindrait seulement 60% (Ponard, 2023). Bien que cette étude concerne la médecine humaine, il n'est pas aberrant de supposer que c'est comparable en médecine vétérinaire. En ce sens, développer un outil capable d'augmenter ce taux d'exactitude diagnostique est d'un réel intérêt.

1.2. Aide à l'interprétation d'examen complémentaire

Dans la continuité de l'aide à la décision diagnostique et toujours en vue de rendre service, des logiciels d'aide à l'interprétation des examens complémentaires grâce à l'IA ont vu le jour en médecine vétérinaire. Ces outils sont déjà nombreux en médecine humaine, notamment dans le domaine de l'imagerie médicale : radiologie, scanner, imagerie par résonance magnétique, échographie, otoscopie.

PicoxIA® permet d'analyser des clichés radiographiques abdominaux, thoraciques et de hanche en extension sur chien, chat et nouveaux animaux de compagnie, afin de donner une interprétation du cliché (Amadieu, 2023). En plus d'être un outil performant et d'avoir pour avantage de ne plus se restreindre aux radiographies thoraciques comme les précédents outils de ce type, il est facile d'accès puisqu'il ne nécessite qu'une connexion internet : il suffit de passer par le site et d'y charger les clichés radiographiques à analyser. Est alors générée une liste de lésions détectées par l'IA avec les degrés de confiance associés (figure 2). De plus, PicoxIA® peut tracer les différents indices de circonstances tel que l'angle de Norberg-Olsson sur des radiographies de bassin, angle formé par la droite reliant les deux têtes fémorales et la tangente à l'angle crânio-acétabulaire tracée depuis le centre de la tête fémorale et qui permet de diagnostiquer une dysplasie des hanches, une pathologie articulaire induisant une mauvaise congruence de l'articulation coxo-fémorale. Cet outil est aussi capable de calculer l'indice de Buchanan qui permet d'évaluer la taille de cœur en comparant différentes mesures réalisées sur une radiographie du thorax. L'IA présente comme avantage d'être objective, ainsi de nombreux biais cognitifs ne sont pas possibles avec cet outil, comme le biais de confirmation signifiant le fait de donner plus ou moins d'importance à certains éléments selon qu'ils soient en faveur ou en défaveur de son hypothèses (Amadieu, 2023). Les vétérinaires n'étant, pour la plupart, pas spécialisés en imagerie ou peu expérimentés, des études réalisées sur quinze type de lésions thoraciques (incluant collapsus trachéal, masse pulmonaire ou encore pneumothorax) ont montré une forte concordance des interprétations radiographiques entre des imageurs spécialisés et des vétérinaires généralistes assistés par l'intelligence artificielle (Boissady *et al.*, 2020). Cependant, il est à noter que ces études ont été réalisées par les développeurs de PicoxIA® ce qui pose ici la question d'impartialité de ces résultats.

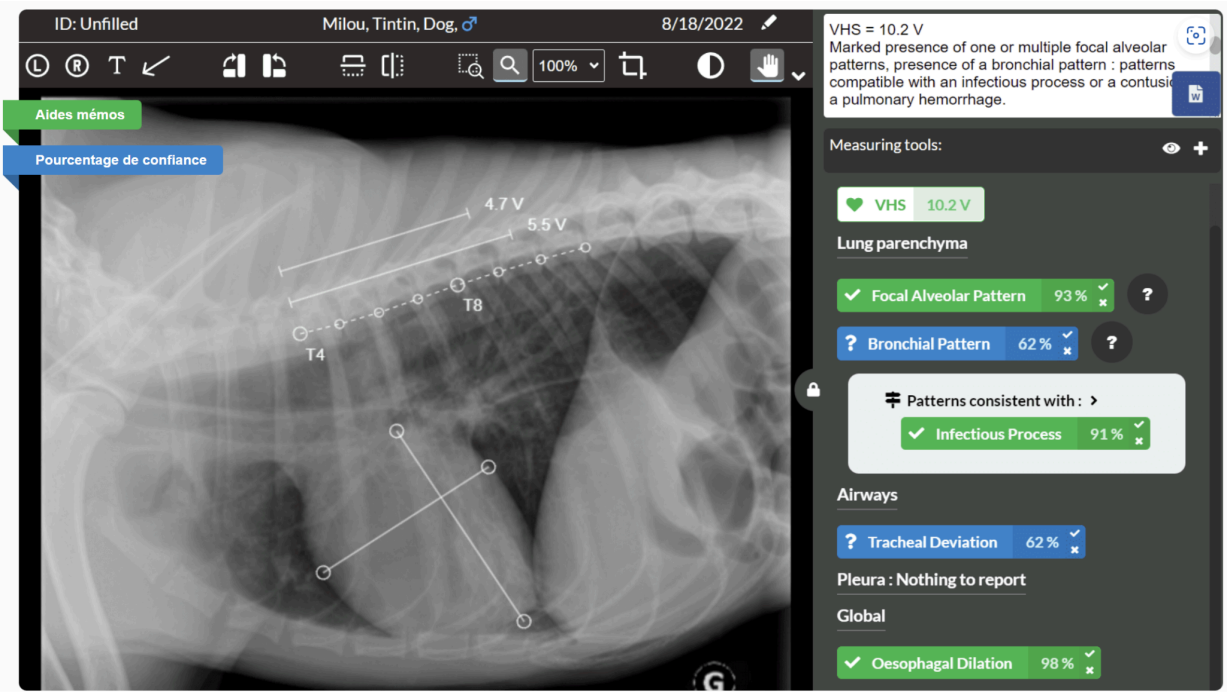


Figure 2 : Exemple d'une radiographie thoracique de chien analysée par le logiciel PicoxIA® montrant différentes lésions observées et leurs degrés de confiance respectifs

Pour d'autres types d'examens complémentaires, Zoetis a aussi lancé sa propre plateforme de diagnostic à partir d'examens complémentaires en utilisant l'IA : IMAGYST®. Cet outil associe un scanner et une technologie d'intelligence artificielle qui analyse les lames de frottis (détection d'une zone de lecture idéale puis numération sanguine), de coprologie (détection des principaux parasites intestinaux du chien et du chat) et des cytologies cutanées avec des résultats comparables à des spécialistes de l'ACVP (American College of Vet Pathologists). L'IA a été entraînée par des spécialistes fournissant des images de haute qualité et leur expertise, sur plusieurs années. De plus, Imagyst® présente l'avantage d'être compatible avec une liste de logiciels de gestion contenant des informations sur les patients (Zoetis France, 2024).

	MACROPHAGES	EOSINOPHILS	LYMPHOCYTES	NEUTROPHILS[1]
Sensitivity (95% CI)	82% (73%-89%)	92% (82%-97%)	87% (80%-93%)	90% (84.5%-94%)
Specificity (95% CI)	88% (81%-92%)	89% (83%-93%)	81% (73%-88%)	92% (84%-97%)

Figure 3 : Précision (en %) de VetScan Imagyst® pour l'identification des différentes cellules inflammatoires sur des échantillons de cytologie cutanée

	COCCI BACTERIA	ROD BACTERIA	*MALASSEZIA*[1]
Sensitivity (95% CI)	77% (69%-83%)	83% (70%-92%)	87% (76%-94%)
Specificity (95% CI)	76% (66%-83%)	73% (66%-79%)	83% (77%-88%)

Figure 4 : Précision (en %) de VetScan Imagyst® pour l'identification d'agents infectieux sur des échantillons de cytologie cutanée.

Vetscan Imagyst® à une sensibilité et une spécificité d'environ 90% pour identifier les cellules de l'inflammation sur des échantillons dermatologiques ainsi qu'une sensibilité et une spécificité avoisinant les 80% pour identifier les agents infectieux lors d'analyses de cytologies cutanées (figures 3 et 4). Ces résultats ont permis de justifier de performances équivalentes à celles de pathologistes spécialisés ACVP cités plus tôt. De même, VetScan analysant l'entièreté de la lame, est exhaustif dans sa caractérisation et a notifié des éléments concernant certaines lames sur lesquelles des pathologistes n'avaient rien détecté (Zoetis US, 2024). Néanmoins, il s'agissait d'éléments qui étaient présents de manière anecdotique et, donc, ayant peu d'influence sur le diagnostic.

De nombreux analyseurs Idexx® utilisés dans beaucoup de cliniques exploitent ce genre de technologie. Typiquement, l'analyseur SédiVUE® utilise l'IA dans l'observation microscopique d'échantillons urinaires, afin de repérer les différents types cellulaires visibles et d'identifier d'éventuels cristaux ou détecter la présence d'agents pathogènes. De même, l'analyseur hématologique ProCyte® utilise l'IA dans l'analyse d'échantillon sanguin, afin d'avoir une meilleure caractérisation de la lignée des globules rouges, de la lignée blanche et des plaquettes, et proposera des outils d'interprétation des résultats obtenus (Idexx, 2024)

1.3. Détection précoce de maladies

L'IA sert d'outil d'aide au diagnostic mais aussi d'outil d'aide à la détection précoce de maladies. En effet, le "machine learning", que nous détaillerons ultérieurement , a pour intérêt de pouvoir gérer les liens entre de multiples variables, même s'ils ne sont pas linéaires. Cela permet ainsi de les tester et de trouver la

combinaison de variables qui a une valeur prédictive intéressante dans la détection précoce d'une maladie.

A titre d'exemple, le maladie rénale chronique touche 15 à 30% des chats de plus de 15 ans et les signes précoces sont frustres et les symptômes ne deviennent évident que lors de phases aiguës ou de stades avancés de la maladie (Masson, 2020). De fait, il a été prouvé que lorsque la créatininémie est augmentée, les reins sont déjà lésés à 75% (Cortadellas, 2024). Or la créatininémie est un marqueur de détection de maladie rénale chronique régulièrement utilisé en clinique, chez le chat.

Ainsi, des outils tels que RenalTech® aux Etats-Unis, ont été développés. RenalTech® a été entraîné sur plusieurs centaines de milliers de données anonymisées afin de pouvoir évaluer la probabilité qu'un chat soit en train de développer une maladie rénale chronique. Plus précisément, cet outil peut détecter l'apparition de cette pathologie, jusqu'à deux ans avant le diagnostic clinique traditionnel avec une précision supérieure à 95%. Afin d'obtenir ce résultat, l'IA combine différentes données, réalisées au cours des vingt-quatre derniers mois, telles que la créatininémie, l'urémie ou encore la densité et le pH urinaire mais aussi l'âge de l'animal puis retourne une valeur indiquant la probabilité que le chat soit en train de développer une maladie rénale chronique. Cela permet ainsi de prédire, dans une certaine mesure, le risque de développement de cette affection et d'agir de façon précoce dans la gestion de la maladie (Mottay, 2023; Lamoly, 2020). Au Japon, Toletta®, une litière connectée et équipée d'un système de reconnaissance faciale est utilisée aux mêmes fins. En effet, par le biais du suivi de nombreuses mesures réalisées telles que le volume d'éjection d'urine, la fréquence d'utilisation et le temps passé dans la litière, l'IA a appris à détecter maladie rénale chronique débutante ou encore d'infection du tractus urinaire (Lamoly, 2020).

Nonobstant le fait que nous présentons surtout l'intérêt de l'utilisation de l'IA en médecine des animaux de compagnie, elle s'est aussi développée dans le domaine des animaux de rente. Bien que son émergence date de 1990, c'est en 2019, que les outils se sont multipliés en vue d'augmenter la performance et la rentabilité des élevages (Lamoly, 2020).

Un premier exemple existe en élevage de volailles : un outil utilisant l'intelligence artificielle a été développé afin de classer les poulets selon leur prédisposition à l'ascite, un épanchement liquidien intra abdominale, en fonction de leur vitesse de prise de poids et des variations de température du bâtiment. Cet outil ayant un taux de réussite dans la prédiction de cette maladie de 100%, elle permet

une économie notable pour les éleveurs de volailles qui peuvent ainsi gérer de façon précoce l'apparition de foyers. Dans l'élevage bovin, l'IA a appris à détecter, entre autres, les mammites subcliniques, notamment en comparant l'analyse des caractéristiques biologiques du lait (conductivité électrique, production, comptage de cellules somatiques…) et les observations comportementales des bovins par l'éleveur ou à l'aide de colliers de détection (Lamoly, 2020).

1.4. La nutrition individualisée

L'IA au service de la santé animale permet aussi de prévenir la prise de poids. La bonne santé d'un animal commence par la gestion du surpoids, très présent chez les animaux de compagnie. L'IA prend place directement au domicile des propriétaires avec, par exemple, les lits équipés de balance qui opèrent un suivi de la variation du poids ainsi qu'un suivi des périodes d'activité et de repos. Ces données, combinées aux informations sur l'alimentation de l'animal fournies par les propriétaires, permettent de détecter précocement des processus arthrosiques ou une apparition de surpoids. Dans le même thème, il existe des jeux robotisés qui vont mesurer l'activité physique de l'animal ou des gamelles intelligentes qui fournissent une ration individualisée pour chaque animal et permettent de mesurer toute variation dans la prise alimentaire.

Les "petfooders" développent directement des applications pour alimenter individuellement les animaux depuis 2019. Par exemple, Individualis® (Royal Canin), est une application à destination des vétérinaires ayant pour but de comparer les caractéristiques physiologiques de l'animal avec ses données médicales afin de choisir l'alimentation la plus adaptée, permettant alors la gestion nutritionnelle la plus optimale possible avec priorisation d'une maladie plutôt qu'une autre dans le cas compliqué d'affections concomitantes. Lorsque la priorisation n'est pas clairement établie comme avec un animal atteint de néphropathie et de diabète, le choix revient alors au vétérinaire (Lamoly, 2020). Cette même étude a montré que, en pratique clinique, uniquement un vétérinaire sur deux propose des calculs de rations et, lorsque cela est réalisé, au moins 50% ne le valorisent pas financièrement et y passent plus de 5 minutes. On comprend alors l'intérêt de la mise en place de ce genre d'application, à plus forte raison dans un contexte où les propriétaires sont de plus en plus demandeurs d'une personnalisation du suivi de leur animal.

Si l'on considère les animaux d'élevage, l'alimentation impacte la santé des animaux mais aussi l'aspect économique de l'entreprise avec son influence sur la fertilité et la production des vaches laitières.

Il existe à ce jour, en élevage aviaire, des algorithmes capables de détecter des carences nutritionnelles, d'analyser les relations entre la consommation de l'aliment, sa qualité nutritionnelle et le poids des volatiles afin d'avoir un système d'alimentation alliant performance et économie. En élevage bovin, des projets d'optimisation de l'alimentation voient le jour avec, notamment, le projet de "mangeoire intelligente" de VikingGenetics®, une entreprise suédoise. Ce projet consiste à utiliser des caméras pour estimer la quantité d'aliments consommés par une vache et de la comparer à sa production laitière et à son poids afin de constater quelles vaches ont l'apport énergétique le plus efficace et le plus rentable (VikingGenetics, 2020).

1.5. Communication animale et analyse des signaux de communication

L''IA se développe aussi peu à peu dans le domaine de la communication animale : c'est tout l'enjeu de "Earth Species Projects", une organisation à but non lucratif, dont le projet est actuellement en développement. Il vise à étendre le travail réalisé avec le langage humain, qui permet d'opérer des traductions d'une langue à une autre et d'avoir une IA capable de saisir la sémantique d'une phrase, pour appliquer cela à tous les types de communication animale. Bien que l'élaboration soit ardue, l'idée est simple : permettre de communiquer entre espèces. Ainsi, cette organisation travaille en synergie avec plusieurs acteurs variés tels que des chercheurs en IA, des biologistes, des entrepreneurs, des enseignants ou même des artistes. A l'heure actuelle, de nombreux travaux de recherche sont réalisées en parallèle : "Earth Species Projects" travaille notamment sur le fait de décoder et traduire certains signaux de communication animale et sur le développement de techniques d'Intelligence Artificielle à même de comprendre une représentation sémantique du langage. Entre autres, certaines études portent sur des expériences de génération de vocalisation, comme le chant des baleines à bosse à l'aide de l'IA, afin de pouvoir tenter de communiquer avec cette espèce (ESP, 2024). Cette organisation n'est pas la seule à s'être lancé ce défi. En effet, on peut retrouver

d’autres travaux sur ce sujet, comme par exemple, celui s'intéressant aux différences entre les chants de l’espèce d’oiseaux *Taeniopygia guttata*, aussi appelé le Diamant mandarin, lorsqu’il sont seuls ou en groupe (Rutz et al., 2023).

On retrouvera toujours la même méthodologie commune à ces expériences. Si l’on prend l’exemple du chant des oiseaux, il s’agit, dans un premier temps, de récolter des données par l’intermédiaire de dispositifs électroniques tels que des traceurs GPS ou des micros enregistreurs, aussi appelé matériel de “biologging”. Ces dispositifs sont utilisés sur un groupe d’oiseaux précis, dans une zone géographique connue et pendant un laps de temps défini. A la fin de la période d’enregistrement, les données récoltées sont soumises à une IA. Cette dernière les associe alors en essayant de trouver une redondance, un schéma, entre les différents paramètres : par exemple, elle va chercher un lien entre un chant précis et un déplacement. Ainsi, l’IA serait capable de reproduire un son spécifique, identifié dans l’exemple à un comportement (figure 5).

La dernière étape consiste alors à expérimenter le modèle de langage créé par l’IA en communiquant directement avec l’espèce étudiée et en observant si elle répond par le comportement attendu ou non.

Ce type de recherche reste encore à développer, l’enjeu éthique sur l’impact du matériel de biologging et des expériences réalisées une fois le modèle de langage mis en place restant à apprécier.

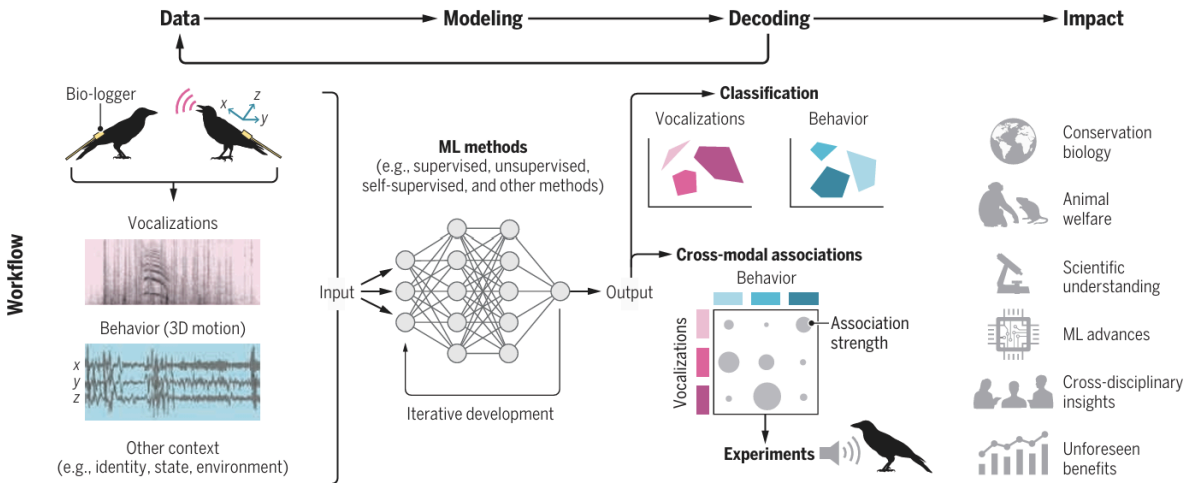


Figure 5 : Schéma représentant la méthodologie d’exploration du langage animal par l’IA

Ainsi, nous venons de voir que l’IA intervient déjà tant auprès du professionnel de santé vétérinaire qu’auprès des propriétaires ou des éleveurs en s’intégrant aux démarches diagnostiques, au suivi des maladies, à l’optimisation de l’alimentation de

chaque individu et permet d'améliorer les performances financières des élevages. A l'avenir, elle pourrait même permettre une communication inter-espèces, une meilleure compréhension des besoins et une optimisation du bien-être animal.

Tout ce qui a été développé ici s'inscrit dans la médecine des 4P qui se veut prédictive, préventive, personnalisée et participative. Ce terme a été breveté en 2013 par l' "Institute for system biology" qui a pour ambition de développer l'idée de "médecine augmentée" (Galland, 2020). Cependant, un domaine reste encore peu concerné par ces avancées : le domaine administratif et plus précisément, la tâche de rédaction de compte-rendu, essentielle et omniprésente dans le quotidien du vétérinaire.

2. Les outils d'optimisation du travail administratif du vétérinaire.

Jusqu'alors, les outils utilisant l'IA pour améliorer le travail et le quotidien du vétérinaire ne concernent que la partie strictement médicale du métier. Or, comme cité en introduction, une partie du temps du vétérinaire est dédiée à la réalisation de comptes-rendus qui permettent le suivi des patients et des actes réalisés. Ce travail est un aspect chronophage du métier de vétérinaire. Nous allons voir que la digitalisation a permis d'apporter du confort sur cet aspect avec des outils variés.

2.1. Logiciels vétérinaires existants et services proposés

Depuis l'ère de la digitalisation, de nombreux logiciels polyvalents de gestion vétérinaire ont vu le jour. Cependant, selon un sondage réalisé en 2007, le marché reste saturé à plus de 60% par 4 d'entre eux : Vétocom®, Bourgelat®, Assistovet® et Vet'Phi® (Médard, 2007). Plus précisément, Vétocom serait le gros vainqueur avec près de 40% des parts du marché (Médard, 2013).

La transition vers le numérique a permis de faciliter la communication avec les propriétaires grâce, notamment, à la prise de rendez-vous par internet ou à l'envoi de messages automatisés un mois avant les rappels de vaccins. Les logiciels de gestion vétérinaire permettent aussi de réaliser les tâches administratives plus facilement et ainsi de se recentrer sur la médecine (Burger, Sigot, 2016).

C’est ce même défi qui a pour vocation d’être relevé dans ce travail : en automatisant au maximum la rédaction de comptes-rendu, l’objectif est de permettre au vétérinaire de se détacher de la prise de note pendant ses échanges avec les propriétaires et ainsi procéder à des interactions plus personnelles avec l’humain et l’animal en consultation. Comme cité plus tôt, la rédaction de comptes-rendus est chronophage et son automatisation représenterait aussi un gain de temps.

En médecine humaine, il existe déjà de nombreux procédés pour diminuer la charge de travail des médecins et du personnel médical en aménageant des solutions d’aide à la rédaction de comptes-rendus telles que Pabau®, un outil de dictée médicale faisant intervenir l’IA avec un vocabulaire spécifiquement scientifique permettant une transcription de ce qui est dit vers un texte le plus fidèle possible (Galloway, 2023). On trouve aussi des sites mettant en relation des professionnels de santé avec des équipes particulièrement formées à la rédaction de comptes-rendus sur la base d'enregistrements fournis par le médecin; ou encore, des logiciels de mise en page de comptes-rendus à partir de réponses à un questionnaire.

En médecine vétérinaire, certaines de ces solutions existent mais sont encore peu démocratisées et peu développées. Parmi les logiciels de gestion cités, seul Vétocom® propose de répondre à ce problème : tout d’abord, il propose l’utilisation de comptes-rendus types qu’il est possible d’appeler dans la zone prévue pour la rédaction afin d’avoir un compte-rendu pré-rempli. Ensuite, il propose une application mobile connectée qui offre la possibilité d’ajouter dans l’historique de l’animal un compte-rendu sous forme de fichier audio. Enfin, Vétocom® va encore plus loin en proposant un outil de diction de comptes-rendus avec reconnaissance vocale (figure 6). Cette nouvelle option est particulièrement intéressante pour les praticiens à domicile.

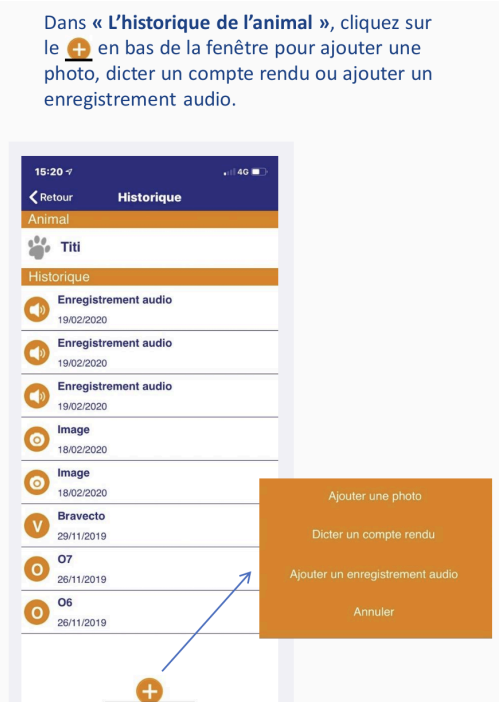


Figure 6 : Interface de l'application mVet de Vétocom® concernant la gestion de l'historique de l'animal et les nouveautés de création de comptes-rendus (Vetocom, 2024)

2.2. Dernières évolutions

Il existe déjà des logiciels de reconnaissance vocale et de diction mais leurs performances sont bien souvent limitées et peu adaptées au vocabulaire spécifique des domaines médicaux. Des projets récents, encore plus avancés, voient actuellement le jour grâce à l'IA.

L'entreprise anglophone Talkatoo®, créée en 2019 à destination des médecins comme des vétérinaires a mis au point une application simple de dictée et de reconnaissance vocale qui, sur presque toutes les zones de traitement de texte d'un écran ou d'un logiciel, est capable de rédiger en temps réel ce qui est enregistré par le micro. Cette application possède une banque de mots adaptés au langage scientifique permettant ainsi une précision très intéressante lors des traductions (Talkatoo, 2024). Malheureusement, elle ne semble pas exister en français pour le moment.

Nuance Communication® est une société spécialisée dans les technologies innovantes utilisant l'IA conversationnelle cognitive et qui présente comme intérêt de proposer ses services dans plusieurs langues mais également dans des secteurs

variés, notamment celui de la santé. Nuance a créé Dragon®, une solution de reconnaissance vocale performante et spécialisée. Il existe ainsi “Dragon Medical One”, une fonctionnalité spécialisée pour les métiers de la santé. Grâce à l’IA, elle permet au professionnel non seulement de faire de la dictée en s’adaptant aux accents, ainsi que de naviguer d’un onglet à l’autre de l’application avec la voix et surtout, elle met à disposition un nouvel outil. En effet, Dragon Medical One propose, aux Etats-Unis et depuis fin 2023, l’outil Dax express qui permet de rédiger automatiquement une ébauche de note à partir de l’échange avec le patient, qui est disponible et modifiable dès la fin de la consultation (Nuance Communications, 2024) (Angevert, 2023). Cette fonctionnalité n’est pas encore disponible autrement qu’en prévisualisation privée en France mais devrait le devenir prochainement. Les tests réalisés aux Etats-Unis montrent un gain de temps d’environ 7 min par consultation. Mais l’entreprise désire approfondir en créant un copilote virtuel d’ici quelques années et ainsi, révolutionner le quotidien du métier des professionnels de santé.

Comme nous venons de le voir, de nombreux outils visant à améliorer le confort et le quotidien des professionnels de santé se développent de plus en plus, en utilisant une IA conversationnelle et cognitive, c'est-à-dire une IA dédiée à la compréhension et la génération de texte combinant des techniques de traitement du langage naturel et d’apprentissage automatique que nous détaillerons ci-après. Plus récemment, Nuance s’est imposée en annonçant Dax express, un nouvel outil ayant la capacité d’initier une prise de note à partir de la conversation entre le médecin et son patient. Cet outil n'est disponible qu’aux Etats-Unis pour le moment mais devrait bientôt arriver en France à destination des médecins. Le présent travail de conception de ReqVet, un outil utilisant une IA conversationnelle et cognitive, a pour objectif d’automatiser la rédaction de comptes-rendus vétérinaires à la fin de chaque consultation et ce, sur la base de l’enregistrement vocal de l’échange entre le vétérinaire et le propriétaire de l’animal. L’idée ici est, non pas d’adapter un outil de médecine humaine à la médecine vétérinaire, mais de créer un produit spécifiquement conçu pour cette dernière.

3. Généralités sur l'intelligence artificielle

Nous allons maintenant développer sur les grands mécanismes de l'IA, notamment de l'IA conversationnelle, afin de permettre d'expliquer ensuite, de façon claire, la mise en place de l'outil ReqVet. Nous verrons que l'IA est basée sur l'algorithmique, une façon de procéder que l'on retrouve dans le cerveau humain. Cependant, pour pouvoir fonctionner, elle nécessite qu'on la "nourrisse" de données afin de réaliser un apprentissage automatique ou "machine learning". Dans cette catégorie on trouve le "deep learning", organisé en réseau de neurones, qui permet de réaliser des tâches complexes. Nous détaillerons ensuite la partie du "machine learning" dédiée à la compréhension et à l'interaction avec le langage humain qui est la catégorie à laquelle appartiennent les outils utilisés pour la réalisation de ReqVet.

3.1. Principe de l'IA

Globalement l'intelligence artificielle regroupe de nombreuses techniques et méthodes visant à reproduire le fonctionnement d'un cerveau humain : " l'IA est fondée sur l'idée que le processus de la pensée humaine peut être mécanisé " (Perrin, 2019). En effet, l'intelligence artificielle se base sur un concept que nous connaissons tous : l'algorithmique. Voici la première partie de la définition d'un algorithme tel que nous l'explique le dictionnaire Larousse : "un algorithme est un ensemble de règles opératoires dont l'application permet de résoudre un problème énoncé au moyen d'un nombre fini d'opérations". Nous utilisons donc sans cesse des algorithmes, par exemple, lorsque nous cuisinons où l'on peut effectivement considérer une recette de cuisine comme un algorithme. En effet, le problème à résoudre s'assimile au plat à cuisiner et l'ensemble fini de règles opératoires correspond aux ingrédients nécessaires et aux étapes constituant la recette afin d'arriver au résultat souhaité. Le cerveau humain lui-même fonctionne parfois de façon algorithmique lorsqu'il effectue une tâche. Il va prendre une suite de décision standardisée afin d'obtenir le plus vite possible et sans erreur la solution au problème posé. Prenons l'exemple du choix d'une pêche mûre sur un étalage, le cerveau va procéder par étape : la pêche a-t-elle bien la bonne couleur? Si oui, a-t-elle une bonne odeur? Si oui, sa texture est-elle tendre? Ainsi, cette suite de petits tests permet de s'orienter sur le choix de la pêche. Chaque individu a ses propres

méthodes pour choisir le meilleur fruit, il existe donc de multiples algorithmes différents permettant d'arriver au même résultat (Lamoly, 2020).

L'IA fonctionne grâce à un panel de données brutes qui permettent l'apprentissage automatique, "machine learning" en anglais.

3.2. "Machine learning" et "deep learning"

Le "machine learning", ou apprentissage automatique, est une science dont l'objectif prioritaire est de fournir des prédictions en se basant sur des statistiques et des données afin de reconnaître des patterns : des répétitions qui peuvent permettre la prédiction. C'est l'utilisation des données passées pour prédire l'avenir (Coheris, 2022). Le plus souvent, on passe par un système d'apprentissage supervisé par un "data scientist", ou expert en science des données, dont la mission principale est de fournir des données exploitables et d'entraîner l'IA. Avec ce système, on fournit en premier lieu et sur une prédiction précise, des données claires X au modèle de "machine learning" et pour lequel le "data scientist" connaît la bonne prédiction Y. Ainsi, l'IA cherchera à estimer la meilleure fonction prédictive f(X) = Y et pourra ensuite appliquer cette modélisation à de nouvelles situations que l'on ne saura pas caractériser (Perrin, 2019). Prenons un exemple concret pour illustrer nos propos : imaginons un modèle ayant pour objectif de reconnaître des fruits sur des photos et pouvoir différencier une pomme d'une orange. Avec le "Machine Learning", il faut définir manuellement des caractéristiques pertinentes comme la forme, la couleur, la taille ou la texture du fruit. Plus précisément, les pommes sont plus lisses que les oranges et leur couleur peut varier du vert au rouge tandis que les oranges sont forcément orange. Grâce à ces informations, sur un nombre de fruits limités, l'IA sera capable de les différencier sur un certain nombre de photos.

Le "Deep learning" est une sous-catégorie de "machine learning" mettant en jeu des réseaux de neurones artificiels. En effet, à l'image d'un réseau neuronal humain, les réseaux de neurones artificiels fonctionnent par couche et plus le nombre de couches est élevé, plus le réseau créé est complexe et précis. Les réseaux de neurones sont capables d'effectuer eux-mêmes des corrections statistiques pour ajuster la prédiction de sortie et ce, sans que le programmateur n'intervienne pour changer les hypothèses et y apporter, par exemple, plus de caractéristiques . Le fait de fonctionner par couche donne l'occasion d'obtenir des

modèles de plus en plus fiables par l'entraînement, chaque couche étant attribuée à l'identification d'une caractéristique en particulier. A chaque nouvelle itération, le "data scientist" valide les bons résultats et infirme les mauvais, permettant ainsi au système de pouvoir modifier les couches neuronales en fonction de ces informations. Si l'on reprend l'exemple des fruits cités plus tôt, avec le "deep learning", il suffit de soumettre à l'IA des milliers de photos de différents fruits, et elle va apprendre seule à différencier les différents types. A force de faire des propositions et de retenir celles que l'opérateur lui aura dit être exactes, le réseau de neurones trouvera, seul, les caractéristiques qui constituent une orange ou un pomme ou encore d'autres fruits. Le "deep learning" nécessite ainsi un grand nombre de données pour s'entraîner mais est aussi capable de produire des résultats plus précis.

Pour résumer, le "machine learning" est donc adapté pour des tâches relativement simples ou lorsque peu de données sont disponibles, alors que le "deep learning" est puissant pour des problèmes plus complexes et des volumes de données massifs, mais demande plus de ressources et d'entraînement.

3.3. "Natural language processing" et "Large Language Model"

Le "Natural Language Processing" (NLP), ou système de traitement du langage naturel, est la partie du "machine learning" s'intéressant à la compréhension et à l'interaction avec le langage humain. Le NLP a de vastes applications comme identifier les sentiments de différents protagonistes dans un texte ou encore intervenir dans le fonctionnement des IA conversationnelles telles que Chat-GPT® (Talbi, 2020). Ce système permet à un modèle d'IA de comprendre, d'analyser et de résumer un texte, par exemple, tout en pouvant commenter l'état émotionnel des différents protagonistes.

Il existe une sous-catégorie du NLP : les "Larges Language Model" (LLM) ou grands modèles de langage, c'est ce qui va nous intéresser ici. A la base, ce sont des modèles statistiques qui prédisent le mot à suivre dans une séquence à partir des mots utilisés précédemment dans celle-ci. A la différence du NLP simple, le LLM est capable de produire un texte cohérent sur la base de données statistiques sans qu'il y ait une réelle compréhension du sens profond du dit texte. Les LLM sont entraînés sur des quantités de données extrêmement importantes et permettent de

rédiger un texte à la manière d'un être humain sans que l'on puisse y déceler une différence.

Afin d'être performant, le LLM utilise également un réseau de neurones qui permet non seulement de traiter les mots mais également les liens entre eux et ainsi de comprendre le sens des phrases. Comme on peut le constater dans la figure 7, les différentes couches de neurones traitent des caractéristiques différentes, étape par étape, pour finalement obtenir des données abstraites sous forme de vecteurs. Précisément, lors de son entraînement, le modèle reçoit des documents en français et en anglais. Il traite chacun des mots dans les deux langues puis il traite les phrases où ces mêmes mots apparaissent, toujours dans les deux langues. A chaque fois, les mots deviennent des vecteurs donc la valeur dépend du mot ou encore de sa place dans la phrase. Que le modèle reçoive le texte en anglais ou en français, il arrivera toujours aux mêmes vecteurs, ce qui permet une traduction dans les deux sens. Cette représentation créée par l'IA, a donc une sémantique commune au texte anglais comme au texte français (figure 7). A partir de cette représentation abstraite, le réseau de neurones peut ainsi, si telle est la demande, retranscrire le texte anglais sous forme de texte français en restant fidèle au contenu de départ et sans opérer une simple traduction mot à mot qui pourrait se révéler inexacte.

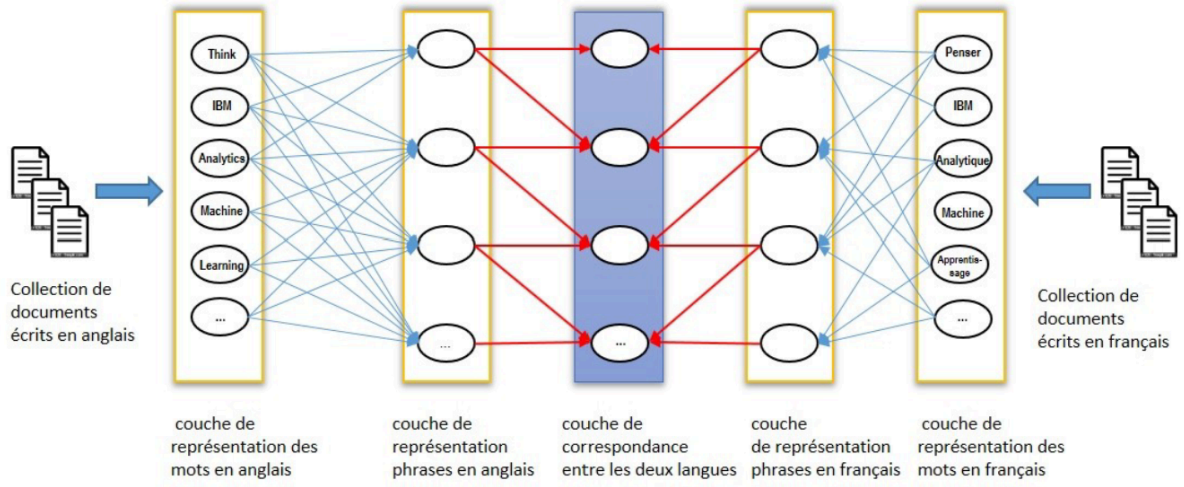


Figure 7 : Illustration d'un modèle de traduction linguistique utilisant un réseau de neurones de type "Natural Language Processing" (NLP) (Perrin, 2019).

Ces modèles reposent sur des architectures de type "transformateur", cela signifie que, comme défini précédemment lors de la définition du "deep learning", ils comportent plusieurs couches de neurones qui ont été pré-entraînés sur des

milliards de pages de textes. Ainsi, cette base de données extrêmement vaste permet que ces modèles d'IA maîtrisent la grammaire, la sémantique et même les représentations conceptuelles ((IBM, 2024).

Ces grands modèles de langage intégrés aux modèles de traitement du langage ont notamment permis la création des IA conversationnelles telles que Chat GPT®. Aussi, il existe désormais des machines qui, non seulement comprennent la sémantique du langage humain mais sont également capables d'y répondre dans un langage sophistiqué et de la même manière que l'aurait fait un être humain. Ce système est notamment utilisé pour les études sur la communication animale et nous l'utiliserons dans cette étude. Plus précisément, le LLM utilisé ici est le modèle GPT ("Generative Pre-trained Transformer"). C'est un des modèles les plus connus qui présente une capacité inégalée à générer des textes cohérents et contextuellement pertinents.

Partie 2 : Création d'un programme de rédaction de compte-rendu à partir d'un fichier audio

Dans le domaine vétérinaire, l'intelligence artificielle est encore peu présente en consultation. En médecine humaine, les outils se développent afin de faciliter la partie administrative du métier avec des systèmes de dictée vocale et est sur le point de connaître une toute nouvelle avancée avec Dragon Medical One cité précédemment. Le but du projet détaillé dans la présente recherche, nommé ReqVet, consiste à intégrer l'IA lors de l'échange entre le propriétaire et le vétérinaire afin de créer un outil capable de rédiger le compte-rendu, ou, a minima une ébauche de ce dernier. L'intérêt est de permettre au vétérinaire de se détacher de la prise de note et d'être ainsi totalement dédié à la conversation durant la consultation. Mais ce ne sont ici pas les seuls enjeux : ce projet représente un réel gain de temps pour le vétérinaire qui n'aurait plus qu'à relire et modifier un compte-rendu final et hiérarchisé. Ce projet a aussi comme objectif de limiter la perte d'informations et de faciliter le suivi des dossiers entre vétérinaires, au sein d'une même structure ou lors de cas référés.

1. Les grandes étapes de l'élaboration du programme ReqVet

La développement de ReqVet s'est articulé en plusieurs segments distincts : en premier lieu, il a fallu passer d'un fichier audio à un fichier texte, c'était l'étape de transcription. A partir du texte obtenu, reflet de la conversation qui a eu lieu lors de la consultation, l'intelligence artificielle devait retourner un compte-rendu vétérinaire clair et exhaustif

La base du projet est donc la rédaction d'un programme informatique à l'aide d'un langage de codage informatique couramment utilisé dans le domaine de l'intelligence artificielle : Python®.

1.1. Cahier des charges

Afin de créer un outil capable de rédiger, de façon standardisé, des comptes-rendus de consultation vétérinaire, plusieurs conditions doivent être respectées.

Premièrement, le programme doit être capable de traiter un fichier audio pour le convertir en fichier texte, appelé transcription, à l'aide d'un système de reconnaissance de la parole ou "automatic speech recognition" (ASR).

Ensuite, à partir du texte obtenu, l'intelligence artificielle opère un triage des informations selon les instructions qui lui sont données puis crée un compte-rendu calqué sur un modèle type de compte-rendu organisé en trois grandes sections : anamnèse - commémoratif - examen clinique. L'IA intégrée au programme doit être capable de rechercher des informations dans un texte et les classer par la suite dans le compte-rendu. Pour obtenir un outil performant, il a été nécessaire de tester plusieurs méthodes faisant intervenir l'IA que nous détaillerons par la suite.

Enfin, l'outil étant basé sur l'enregistrement de conversation et dans le but de respecter le secret professionnel, il est nécessaire de protéger les données utilisées. Pour ce faire, l'utilisation d'outils en accord avec les normes RGPD et le secret professionnel était inévitable. Les deux étapes du programme étaient concernées : l'enregistrement vocal et sa transcription, ainsi que le compte-rendu généré à partir de ces fichiers. Ainsi, chacun des outils qui nous détaillerons respectent donc les règles de protection et de confidentialité des données.

De plus, le programme a été rédigé sur Google Colab®, un service gratuit proposé par Google® qui est un équivalent de Google Drive® mais adapté au codage informatique. Google Colab® permet de coder en ayant accès à des ressources suffisantes pour faire fonctionner le programme informatique.

1.2. "Speech-to-text"

Le premier modèle d'IA que nous avons utilisé permet de convertir l'enregistrement audio de la consultation en un texte. En effet, afin que l'algorithme puisse traiter cet enregistrement, il est nécessaire de passer par un type de fichier qu'il est capable de traiter, en l'occurrence en un fichier texte, également appelé une chaîne de caractère en langage Python®.

Pour ce faire, c’est l’outil Whisper®, un système de reconnaissance automatique de la parole, qui a été choisi pour réaliser cette action. Whisper® a été créé par la société OpenAI®, également créateurs de ChatGPT, l’intelligence artificielle conversationnelle.

L'intérêt de Whisper® est multiple : en effet, il a été entraîné avec plusieurs langues et langages techniques mais aussi sur des voix avec des accents différents et des enregistrements avec des bruits de fond. De même, Whisper® transcrit le texte en ajoutant la ponctuation et sait différencier les voix des différents protagonistes. C’est donc un outil extrêmement performant. Enfin, il est, comme tous les outils utilisés pour la réalisation de ce projet, disponible en format Open Source. C’est-à-dire qu’il est accessible sous licence libre depuis le 21 septembre 2022 (Jeanviet, 2023).

Whisper® permet de transcrire un enregistrement vocal en texte dans de nombreuses langues ainsi que de le traduire en anglais ou de traduire de l’anglais vers la langue souhaitée. Pour ce faire, OpenAI® a entraîné Whisper® sur 680 000 heures d’audios associé à la transcription correspondante dont 17% des fichiers étaient des audios dans une autre langue que l’anglais avec leur transcription associée dans cette même langue. Il a été établi que la performance de Whisper® dans un langage donné est directement corrélée au nombre de données utilisées pour l'entraîner. Or, parmi ces 17% de données, l’adaptation de Whisper® au langage français a été effectué à l’aide de 9 752 heures d’audio au total, classant cette langue dans le Top 5 des langues les plus représentées dans son entraînement à la transcription (Annexe 1), ce qui en fait un excellent outil pour la reconnaissance vocale française.

De plus, il existe différentes tailles de modèle Whisper® (figure 8) à adapter selon l'objectif et la puissance de calcul disponible. Plus la taille du modèle est importante, plus le modèle est performant et précis. Cependant, il demande aussi une quantité de ressources plus importante pour fonctionner. Il s’avère, comme expliqué précédemment, que par l’intermédiaire de Google Colab, nous avons pu utiliser le modèle le plus complet et donc, le plus performant dont la dernière version mise à jour date de Novembre 2023 (GitHub, 2024).

Size	Parameters	English-only model	Multilingual model
tiny	39 M	✓	✓
base	74 M	✓	✓
small	244 M	✓	✓
medium	769 M	✓	✓
large	1550 M		✓

Figure 8 : Tableau récapitulatif des différentes tailles de modèles Whisper® existants

Whisper® est constitué de sorte à traiter des segments d'audios de 30 secondes sur des enregistrements de quelques minutes à quelques heures. En effet, l'enregistrement est divisé en fractions de 30 secondes qui sont traitées consécutivement et chacun de ces extraits chevauche le précédent afin de conserver le sens lorsque les transcriptions sont fusionnées. Afin d'évaluer la performance de Whisper® sur des transcriptions de longue durée, Radford *et al.* ont exploités sept ensembles d'enregistrements anglais de longueurs et de conditions d'enregistrement différentes, allant de TED talk - conférences de 18 min maximum sur une grande variété de sujets - à des interviews intégrales dans un anglais variant selon les régions, en passant par des extraits de l'émission the Late Show présentée par Stephen Colbert, chargés de jargon et d'expressions du langage commun. En plus de Whisper®, quatre autres modèles commerciaux de système de reconnaissance automatique de la parole et un modèle open-source (NVIDIA STT) ont été testés sur leur performance de transcription. Les différents taux d'erreurs sur les mots (WER) ont alors été comparés. Cette étude a montré que Whisper® surpasse le modèle open-source, sur tous les ensembles de données, et dans la plupart des cas, surpasse également les systèmes ASR commerciaux. (Radford *et al.*, 2022) (figure 9)

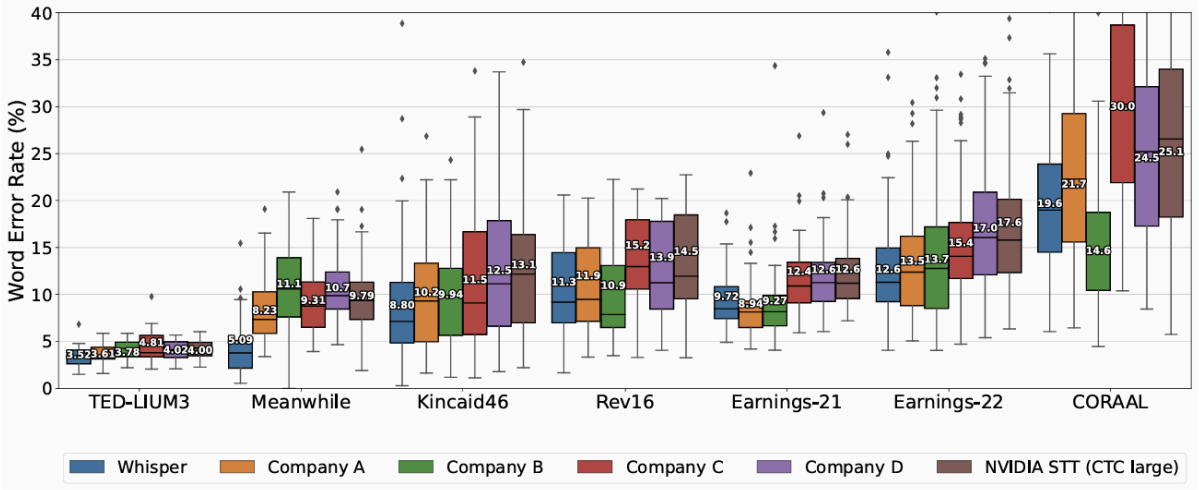


Figure 9 : Comparaison du taux d'erreur sur les mots entre Whisper®, 4 modèles commerciaux de système de reconnaissance automatique de la parole et NVIDIA®, un autre ASR Open-source

D'autres tests montrent que, bien que Whisper® n'atteigne pas une performance parfaite, elle est néanmoins similaire à la précision humaine (Radford et al., 2022).

Il est à noter que ces tests ont été effectués sans avoir entraîné précisément Whisper® à une banque de mots spécifiques mais, au contraire, en conservant la forme en open-source afin de connaître ses performances dans les cas les plus généraux possibles. C'est pour cette raison que cette étude est pertinente: en effet, dans notre cas, Whisper® n'a pour le moment, jamais été entraîné au préalable sur des données liées aux comptes-rendus et au vocabulaire vétérinaire.

1.3. Choix du LLM

Une fois le texte obtenu, à partir de l'enregistrement, il a fallu choisir un LLM ou grand modèle de langage qui, comme expliqué en première partie, peut comprendre le langage humain et répondre avec un langage adapté, semblable à la façon dont l'aurait fait un être humain. Les LLM étant en constante évolution, nos choix ont changé au cours du temps et ont été modifiés tout au long du développement du projet.

Lors des premiers essais, notre choix s'était porté sur Mistral 7B®, un LLM créé par une entreprise française, Mistral AI, comprenant 7 milliards de paramètres.

Or, comme pour Whisper®, la taille du modèle conditionne son efficacité. En effet, les paramètres sont des valeurs que le modèle “apprend” à partir des données de son entraînement et plus un modèle a de paramètres, plus il peut capturer des nuances et des complexités dans le langage. Cependant, plus le nombre de paramètres est élevé, plus le modèle nécessite de puissance de calcul pour pouvoir fonctionner.

En effet, au moment de notre premier choix, plusieurs LLM en Open Source ont été testés. Les plus connus et les plus utilisés étant LLama 2 13B® de Meta et Mistral 7B® de Mistral AI : ce dernier se démarque cependant dans la littérature non seulement par des performances excellentes sur un large éventail d’applications, mais également par le fait qu’il est conçu avec un nombre de paramètres relativement faible ce qui permet de l’utiliser plus facilement et rapidement sans compromettre les performances du LLM. De plus, Mistral 7B® a démontré avoir moins tendance aux hallucinations, inconvénient retrouvé dans les LLM et lié à l’aspect prédictif de leur fonctionnement : les LLM utilisant des statistiques pour prédire le mot suivant dans une génération de phrase, il arrive que le contexte ne soit pas pris en compte et que la génération d’informations soit factuellement incorrecte ou non pertinente (Elias, 2024). Mistral 7B® est donc plus fiable dans ses résultats. En effet, Mistral s’avère plus performant que presque toutes les versions de Llama® en terme de raisonnement, de compréhension et de connaissances : on remarque qu’à nombre de paramètres égal, et donc à vitesse d’exécution égale, il dépasse largement le LLM de Meta. On peut même ajouter que Mistral 7B® surpasse également la version de Llama à 13 milliards de paramètres (figure 10). Ainsi, Mistral 7B® semblait être le meilleur compromis : peu de paramètres pour une performance conservée, ainsi, il nécessite une ressource minimale pour pouvoir fonctionner.

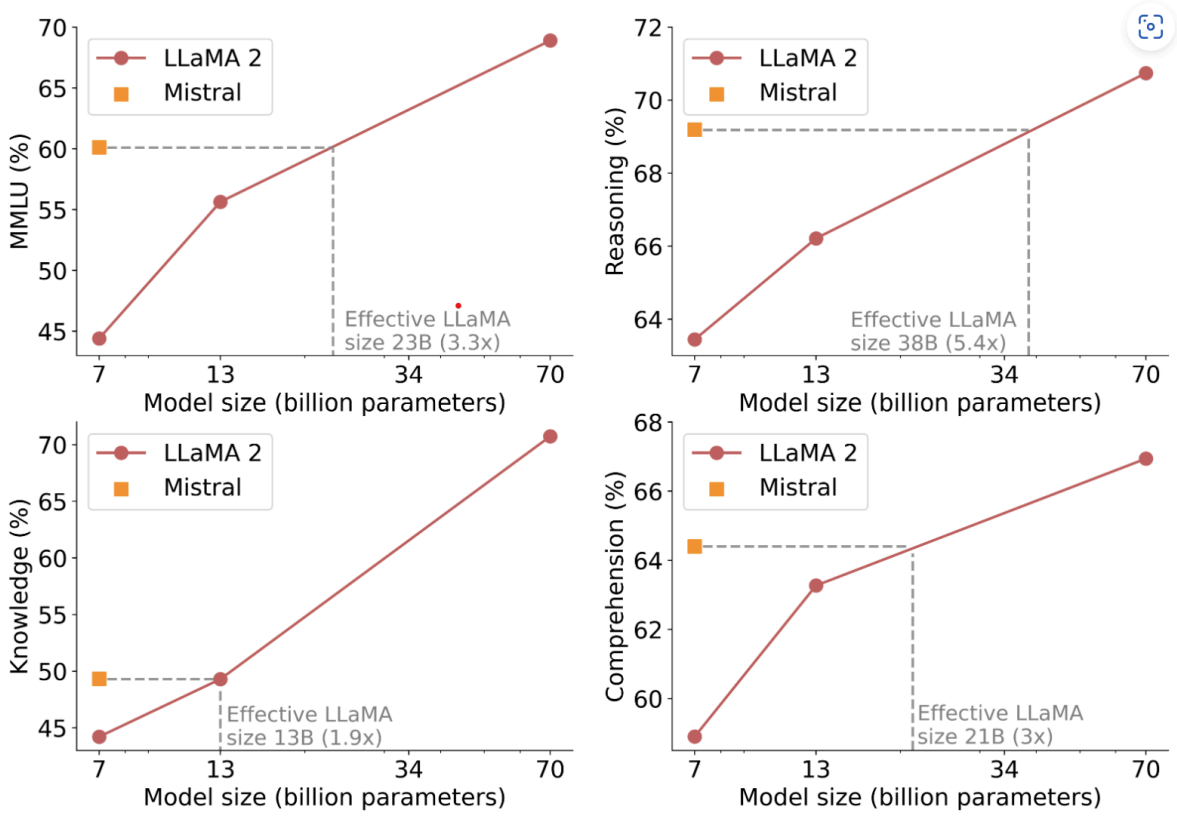


Figure 10 : Comparaison des performances de Mistral 7B® avec Llama 2® (7B/13B/70B) (MistralAI, 2023)

1.4. Utilisation de la méthode RAG

La méthode RAG, “Retriever-Augmented Generation” ou Génération augmentée par récupération, est une technique qui permet d'améliorer les performances des modèles de langage comme les LLM évoqué précédemment en réduisant les erreurs d'« hallucination » et d'extrapolation, où le modèle pourrait inventer des informations (Prompt Engineering Guide, 2024a). Pour cela, la méthode RAG combine les connaissances du LLM avec une base de données externes (Gao et al., 2024) : comme le montre la figure 11, l’utilisateur pose une question, ici appelée requête, de type “quel est l’impact du réchauffement climatique sur les océans?”. Le système RAG ne se contente pas de chercher une réponse uniquement dans les informations que le modèle a apprises lors de son entraînement mais va aussi comparer cette question avec une base de données externe contenant des documents ou informations spécifiques qui ont été préparés à l'avance et soumises à l’algorithme.

Cet index de document, pour pouvoir être analysé, va être découpé en petits segments et chaque segment subira un processus appelé “embedding” : c’est-à-dire qu’il va être transformé en vecteur. Un vecteur est une manière de représenter les mots ou phrases sous forme de nombres, afin que l'ordinateur puisse les analyser plus facilement. Ensuite, chacun de ces vecteurs est stocké dans une base de données.

Quand l'algorithme reçoit une question, il convertit de la même manière cette question en vecteur et effectue une recherche pour trouver les segments de la base de données qui sont les plus similaires à la question.

Ainsi, avec le RAG, l’algorithme va convertir les données externes fournies au départ en une base de données de vecteurs chiffrés avant d’effectuer une recherche de similarité entre les vecteurs de la base de données et les vecteurs qui représentent la requête imposée au départ. Une fois qu'il a trouvé les informations les plus pertinentes, l'algorithme combine ces informations avec sa propre connaissance pour générer une réponse (figure 11).

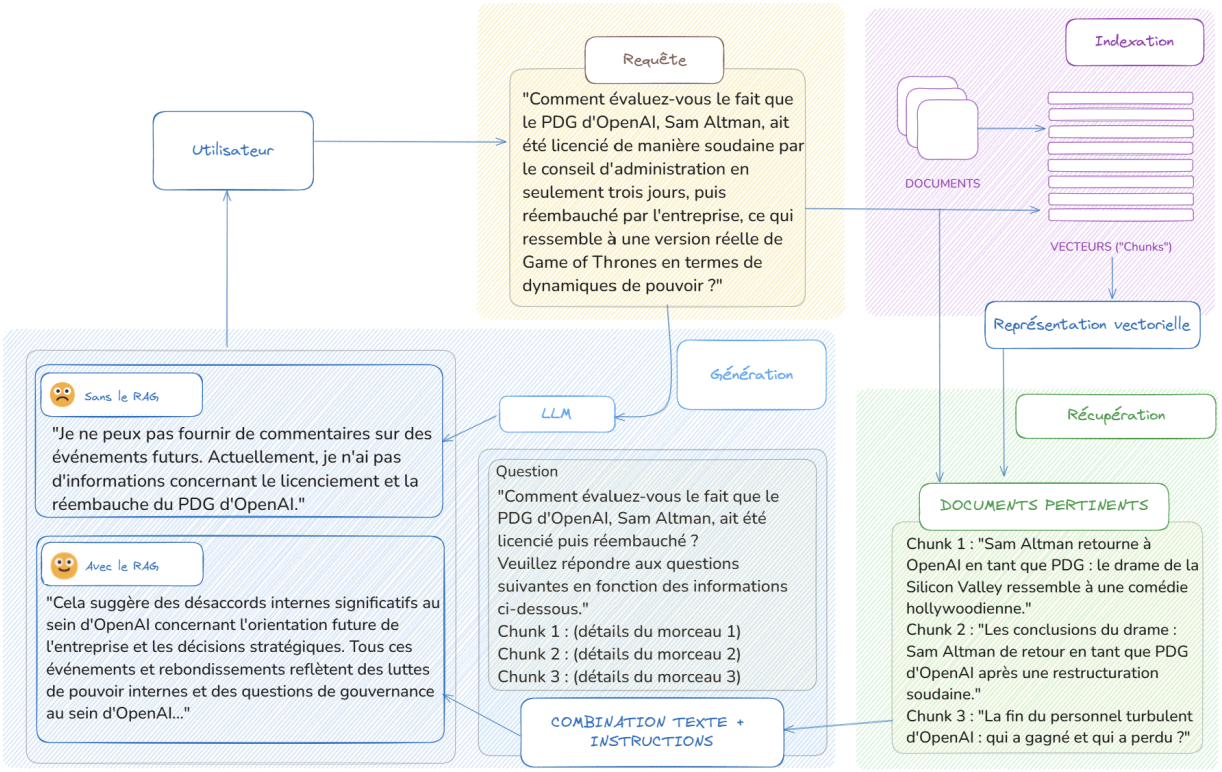


Figure 11 : Traduction française du schéma explicatif du RAG dans l’étude de Gao et al. (Gao et al., 2024)

Dans notre travail, la méthode RAG a été appliquée avec, comme données externes, l'enregistrement de la consultation, créant ainsi à chaque itération de l'algorithme une nouvelle base de données spécifiques à chaque consultation. Ainsi, les requêtes étaient posées une à une comme nous le verrons ensuite et le RAG permettait une réponse uniquement basée sur la conversation avec le propriétaire, limitant ainsi le risque d'hallucination (Figure 12).

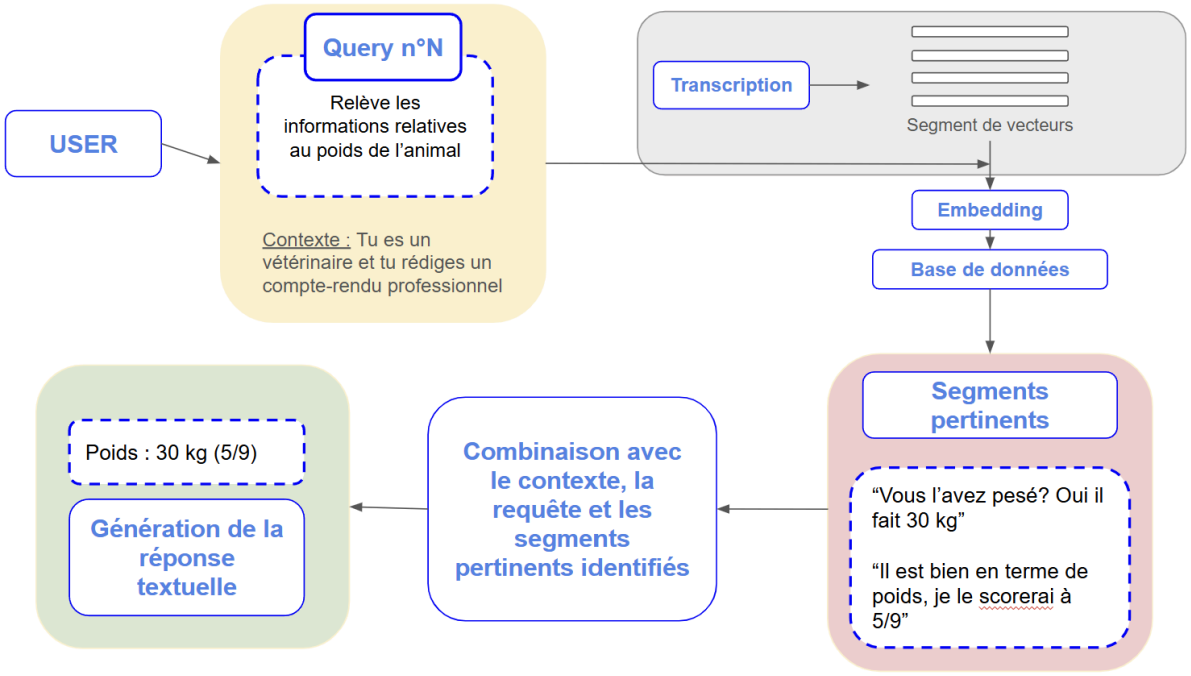


Figure 12 : Schéma explicatif du fonctionnement de ReqVet, basé sur l'explication du système de génération augmentée par récupération

1.5. Deuxième version de ReqVet : Utilisation des "transformers"

Après quelques essais avec Mistral 7B, les résultats étaient prometteurs sans être satisfaisants, car encore trop aléatoires.

Jusqu'alors, nous utilisions le LLM sur notre disque dur local, en téléchargeant ce dernier et ce, afin de pouvoir faire tourner l'algorithme sans réseau internet et afin de conserver les données confidentielles sur le disque dur local. Cependant, cela imposait de réduire le poids du LLM avec le principe de la quantification, qui consiste en une conversion au sein d'un LLM d'une "représentation de données de haute précision en une représentation de données de moindre précision, c'est-à-dire d'un type de données qui peut contenir plus d'informations à un type qui en contient moins" (Talamadupula, 2024). Plus simplement, on peut illustrer la quantification du

LLM de la manière suivante : c'est comme si on réduisait la taille d'un gros dessin en utilisant moins de couleurs, pour qu'il prenne moins de place tout en gardant l'idée de base. Ici, la quantification a donc permis de diminuer le poids de Mistral 7B au détriment de sa précision, d'où des réponses trop approximatives de l'algorithme à nos demandes.

Ainsi, nous avons décidé d'utiliser un LLM qui n'était pas Open Source afin de pouvoir constater si les résultats obtenus étaient meilleurs. Pour ce faire, notre choix s'est porté sur les LLM de type GPT ou "Generative Pre-trained Transformer", or les "transformers" ont la particularité de pouvoir être entraînés sur de grandes bases de données et ce, très rapidement. Ces modèles sont basés sur 3 principes illustrés sur la figure 13 : l'encodage de position qui améliore les prédictions car le modèle comprend l'ordre des mots en leur attribuant une valeur chiffrée en fonction de leur place dans la séquence ; le principe d'attention faisant que le modèle se concentre sur les mots les plus importants dans une séquence ce qui permet la prise en compte de la dépendance grammaticale forte entre ceux-ci, surtout en français ; ainsi que le self-attention permettant une pondération différente des mots en fonction de ceux auxquels ils sont rattachés dans la séquence et donc, une meilleure compréhension du contexte global. L'exemple utilisé ici est la phrase "Le chat noir mange le poisson". L'encodage de position attribue un chiffre à chaque mot en fonction de sa position dans la phrase. Le principe d'attention permet au modèle de se concentrer sur certaines parties de la phrase : en analysant le mot "mange", une plus grande attention sera prêtée aux mots "chat" et "poisson", car ils sont liés au sujet et à l'objet de l'action (Schiltz, 2023; Talbi, 2023)(Figure 13). Enfin, le self-attention évalue les relations entre les mots pour permettre la compréhension du contexte. Les "transformers" comme GPT, bien que non Open-Source, sont donc fortement intéressants dans la réalisation de notre algorithme et c'est pourquoi notre choix s'est porté sur le LLM d'OpenAI®.

Figure 13 : Principe de fonctionnement des "transformers".

Nous avons donc utilisé la version payante de GPT-3.5 puis GPT-4, un LLM à 175 milliards de paramètres, ce qui en fait un modèle nécessitant une puissance de calcul très importante. La différence, en plus du nombre de paramètres et de la meilleure fiabilité du modèle, réside dans le fait que le LLM n'était plus téléchargé sur le disque local mais que nous passions maintenant par la société qui a créé ce LLM, OpenAI®.

Plus précisément, chaque demande représente une requête qui est envoyée sur le serveur d'OpenAI® nous renvoyant leur réponse de manière chiffrée après avoir interrogé le LLM. De ce fait, nous avons pu mettre de côté le compromis sur la performance de l'algorithme en trouvant une solution qui permet à GPT-4 de fonctionner à pleine capacité, tout en protégeant les données selon les normes RGPD, comme expliqué précédemment. Il est à noter que ce LLM à 175 milliards de paramètres n'aurait pas pu fonctionner avec la puissance d'un ordinateur de bureau. En effet, pour accéder à GPT-4, il est nécessaire de passer par le serveur de la société OpenAI®. En effet, payer un abonnement mensuel permet notamment d'obtenir une sécurisation et une protection des données en plus d'utiliser GTP-4 sans souci de ressources.

Les premiers essais ont été réalisés avec GPT-3.5 Turbo, qui était la dernière mise à jour disponible avant la mise à disposition de GPT-4 Turbo en novembre 2023. A la sortie de ce dernier, le programme a été mis à jour afin d'utiliser le dernier modèle sorti, légèrement plus performant. Cependant, ces deux modèles de LLM sont, à la base, des IA conversationnelles, ainsi elles sont optimisées pour réaliser

des échanges et donc, connaissent des limites à la recherche d'informations dans un texte comme ce qui est attendu avec notre algorithme. Enfin, OpenAI® a donné accès à GPT-4o en mai 2024, deux fois plus rapide et deux fois moins cher que ses prédécesseurs (OpenAI, 2024). Il s'est avéré être aussi bien plus efficace dans l'interprétation de texte, comme nous le verrons ci-après.

1.6. "Prompting" : L'art de la formulation

Comme expliqué en partie C avec le schéma d'illustration du RAG, celui-ci débute avec une "query" ou requête pour cibler sa recherche. Ainsi, la trame de l'algorithme est une somme de vingt-six requêtes, soumises aux systèmes RAG concernant chaque étape du compte-rendu, appelé des prompts. Ces requêtes sont aussi celles soumises au LLM de type "transformers" dans la deuxième version de ReqVet. L'élaboration de ces prompts a été une étape longue, nécessitant de nombreux essais afin de trouver la meilleure formulation menant à un résultat correct et répétable.

Dans notre cas et avec les deux versions, c'est-à-dire celle utilisant un système RAG et celle faisant usage de "transformers", les prompts comprennent toujours les mêmes éléments : d'abord, un contexte est donné afin d'orienter le modèle vers un vocabulaire scientifique mais aussi vers un mode d'expression rédactionnel et officiel. Une fois le contexte posé, une question - ou requête - y est associée.

En "prompting", la qualité de la réponse dépend non seulement de la qualité du prompt - de sa formulation - mais également de sa précision. Il existe de nombreuses techniques de prompting, un exemple étant la méthode RAG détaillée précédemment. La méthode utilisée dans la version de ReqVet avec les "transformers" est légèrement différente. L'application d'un LLM de type "transformers" a entraîné l'utilisation d'une méthode inspirée du prompt chaining afin d'augmenter les performances de ce dernier (Prompt Engineering Guide, 2024b). En effet, la recherche d'informations dans un texte est une tâche qui nécessite des prompts détaillés avec de nombreuses précisions. Or, une longueur trop importante de prompt peut être délétère et la réponse à la requête incomplète. Avec le prompt chaining, la tâche principale est détaillée en sous-tâches. Ainsi, les différentes

sous-tâches sont effectuées puis intégrées dans le prompt plus large qui comprend le contexte global et la structure du compte-rendu final.

1.7. Mise en forme du compte-rendu

Une fois les informations pour chaque requête collectées et stockées dans les différentes variables correspondantes, celles-ci sont intégrées dans une chaîne de caractères correspondant à l'articulation générale du compte-rendu de consultation de médecine classique.

Le modèle de compte-rendu utilisé ici est basé sur les modèles de comptes-rendus en médecine préventive et en médecine générale à disposition sur le logiciel du Centre Hospitalier Universitaire Vétérinaire de l'Ecole Nationale Vétérinaire de Toulouse, Sirius® (Annexe 2). On y retrouve donc, dans l'ordre, les informations suivantes : commémoratifs, anamnèse et examen clinique. En première partie, les commémoratifs rassemblent l'ensemble des informations sur le passé clinique de l'animal mais aussi sur sa vie au sens large telles que le statut physiologique, son alimentation, les traitements antiparasitaires, le statut vaccinal, etc. Dans un second temps, l'anamnèse qui correspond à l'historique de la maladie actuelle du patient est détaillée. Elle est centrée sur le motif de consultation et regroupe des informations telles que les signes cliniques observés, leur évolution, les éventuels examens déjà réalisés, les éventuels traitements… Enfin, est rapporté l'examen clinique concis et complet. Ainsi, le programme informatique retourne à chaque fois un compte-rendu avec cette même structure mais personnalisé avec les informations recueillies lors de la consultation (Annexe 3).

1.8. Normes RGPD et sécurisation des données

L'aspect de sécurisation des données a été un point non négligeable dans la mise en place de nos choix. En effet, le vétérinaire, comme plusieurs autres professionnels, est tenu au secret professionnel. Le non-respect de cette obligation est passible de sanction disciplinaire voire pénale, pouvant entraîner jusqu'à un an d'emprisonnement et 15 000€ d'amende (Kieffer, 2014). Par conséquent, les locaux et le matériel informatique, qui permet le stockage et la communication des données, doivent être adaptés à la mise en respect du secret professionnel. De ce fait, dans

l'élaboration de ReqVet, les enregistrements et les comptes-rendus sont inclus dans les données soumises au secret professionnel. Bien que les enregistrements effectués ne fassent pas état de données personnelles sensibles mais simplement de données factuelles concernant l'animal, il a tout de même été de mise d'être attentif sur deux points : la transcription de l'audio en vocal par Whisper® et la génération du compte-rendu à partir de ladite transcription.

Ainsi, la tenue du secret professionnel repose également sur le respect des normes RGPD, ou règlement général de protection des données, de nos différents outils (Hervey-Chupin, 2022).

Concernant la première partie, les phases de tests ont été effectuées en utilisant le modèle Whisper® accessible par Google Colab®. Cet outil ne sera pas conservé pour le futur développement de ReqVet : certains serveurs ne sont pas européens et ne sont donc pas soumis au respect des normes RGPD. Ainsi, malgré le fait que nous rémunérons Google® pour avoir accès aux ressources nécessaires pour faire tourner notre programme, nous ne pouvons que soupçonner le respect de la sécurité des données sans pour autant en être sur. C'est pourquoi, lors des phases de test présentées ici, les enregistrements vocaux étaient anonymisés et ne contiennent aucunes données sensibles.

A l'avenir, l'utilisation et l'accès à Whisper® se fera par le biais de la plateforme RunPod.io® qui héberge, elle aussi, un modèle Whisper®. Or, cette plateforme nous permet de l'utiliser tout en étant assurés de respecter les normes RGPD en passant par des centres de traitement des données européens, respectant toutes les normes de sécurisation des données. De plus, le serveur Runpod® lui-même cherche activement à obtenir ces mêmes certifications d'ici fin 2025, d'après les échanges réalisés avec eux par mail (Runpod.io, 2024) (Annexe 4)

Pour ce qui est de la génération même du compte-rendu, notre premier choix s'était porté sur la méthode RAG qui, bien que moins performante, permettait alors l'utilisation de LLM open source en optimisant les résultats obtenus comme expliqué au début de cette étude. L'avantage de l'utilisation de modèles open-source était justement que nous pouvions héberger ce modèle localement et ainsi, effectuer l'opération sur le disque dur local sans envoyer les données en réseau, le secret professionnel était alors respecté sans ambiguïté. La sécurisation était optimale mais, comme l'ont montré les résultats, le compte-rendu généré était peu performant avec une forte variabilité de la qualité de la réponse et, en moyenne, seulement 70% du contenu était exact.

Ensuite, la société OpenAI® a permis l'accès aux "transformers" avec un abonnement, payant mais peu onéreux, en passant par leur plateforme tout en respectant également les normes RGPD. Un portail internet renseignant toutes les informations relatives à la sécurité et aux données est à disposition des utilisateurs. Il y est notamment précisé que les données au repos et en transit sont chiffrées mais également que OpenAI® est conforme à plusieurs types de règlement dont les normes RGPD (OpenAI, 2024). De plus, il est établie que l'abonnement souscrit à OpenAI® pour l'utilisation des "transformers" inclut un DPA (Data Processing Agreement), soit un accord entre le responsable des données (ici ReqVet) et le sous-traitant (OpenAI®), qui assure la conformité du traitement de la donnée en accord avec les normes RGPD ou autres. Les données ne sont pas sauvegardées, et sont chiffrées via le standard durant leur transit.

Partie 3 : Test pratique et évaluation de l'efficacité de l'algorithme

1. Matériel et méthodes

Les enregistrements vocaux ont été collectés lors des pré-consultations de médecine préventive, de médecine générale, de médecine interne, de reproduction et également d'urgences et soins intensifs au Centre Hospitalier Vétérinaire des animaux de compagnie de l'Ecole Nationale Vétérinaire de Toulouse. Le choix a été porté sur les pré-consultations afin de pouvoir enregistrer un échange le plus proche possible des échanges réalisés en clinique privée entre le vétérinaire et les propriétaires. De plus, dans les critères d'inclusion, figurait également le fait que nos choix se soient portés sur des individus venant à l'ENVT pour la première fois ou, du moins, ayant un intervalle de temps important entre le jour de la consultation et la dernière visite afin que la prise de commémoratifs et le résumé des antécédents médicaux soient les plus exhaustifs possibles. Par exemple, lors des consultations de médecine préventive, ont été exclues les consultations où les personnes amenant leur animal pour deuxième ou troisième injection de primo-vaccination. En effet, ces derniers étant venus pour la dernière fois il y a environ un mois, ces échanges n'étaient pas pertinents pour l'évaluation du projet puisque toutes les questions portant sur les commémoratifs n'étaient alors pas reposées aux propriétaires.

Les enregistrements en eux-mêmes ont été réalisés à l'aide d'un téléphone portable de type smartphone, posé sur la table de consultation à côté des protagonistes de l'échange. Ils comprennent la prise d'informations avec les propriétaires, c'est-à-dire les commémoratifs et l'anamnèse, ainsi que l'enregistrement de l'examen clinique prononcé à voix haute, sans ordre prédéfini d'énonciation.

Afin de pouvoir réaliser ces enregistrements, un consentement éclairé a été présenté et signé par les propriétaires avant le début de chaque enregistrement afin d'obtenir leur accord pour l'enregistrement et l'utilisation des données (Annexe 5).

2. Modalités d'évaluation du programme

Pour évaluer la caractérisation et le tri d'informations réalisé par l'algorithme, organisés en une suite de vingt-six requêtes, nous avons noté chacune des vingt-six requêtes par une notation allant de 0 à 2 (tableau 1).

Tableau 1 : Tableau explicatif du système de notation mis en place

Note	Signification
0	Réponse fausse
1	Réponse incomplète - Réponse incohérente - Réponse complète mais associée à un bug (cf. ci-après)
2	Réponse complète ou réponse avec un léger manque de précision mais comprenant toutes les informations globales

L'addition de toutes les notes attribuées à chaque requête a été effectuée pour obtenir une note globale pour chacun des comptes-rendus rédigés avec l'algorithme. La note maximale possible sur un compte-rendu est donc de cinquante-deux, en effet cela représente le cas où chaque requête a correctement été respectée et donc où les vingt-six requêtes ont eu comme attribution la note de deux.

On peut ensuite obtenir une note moyenne reflétant la précision globale de l'algorithme et celle de chaque requête individuellement afin de d'observer si une des requêtes en particulier est moins performante que les autres et pourrait être à l'origine d'une diminution de la performance globale.

Les écarts-types ont aussi été calculés, permettant ainsi d'évaluer la variabilité du programme et sa fiabilité : un écart-type élevé est donc signe d'une fiabilité amoindrie et d'une répétabilité du résultat peu élevée.

Chacun des calculs a été réalisé avec Google Sheet, l'équivalent d'Excel.

Chacun des enregistrements a été testé sur la première version basée sur l'utilisation du "Retriever Augmented Generation" et sur la deuxième utilisant les "transformers" afin d'en comparer l'efficacité. Les essais réalisés avec la deuxième version ont été effectués avec les trois modèles de LLM cités précédemment : GPT-3.5 Turbo, 4 Turbo et 4o.

3. Résultats

Un total de treize pré-consultations a été enregistré dont quatre de médecine préventive, quatre en consultation d'urgences, trois de médecine interne, une de médecine générale et une de reproduction. La prévalence des différents services dans ces tests est corrélée à la densité des consultations accueillies à l'ENVT et à la prévalence des différentes rotations dans le planning des étudiants de dernière année : la médecine interne et les urgences représentant la majorité de celles-ci et la médecine préventive, quant à elle, présentant un nombre important de consultations journalières. A ces treize enregistrements sont venues s'ajouter trois consultations fictives, réalisées au début de la rédaction du programme, mettant en scène des rendez-vous de médecine générale vétérinaire avec des protagonistes toujours différents.

Cette forte diversité de provenance des consultations a permis d'observer l'efficacité du programme dans plusieurs situations : en effet, certains enregistrements fictifs présentaient au final peu d'informations, d'autres font état de bruits de fond importants ou de plaintes d'animaux. D'autres encore, mettent en scène des propriétaires avec un fort accent et une maîtrise approximative de la langue française. Ainsi, bien que le nombre d'enregistrements soit faible, les conditions ont pu être variées et ont permis d'augmenter la valeur de notre test.

Les seize enregistrements ont alors été testés quatre fois : une fois sur la première version de ReqVet puis, comme expliqué précédemment, trois fois sur la deuxième version avec les trois générations de "transformers" qui sont, pour rappel, des modèles de traitement du langage permettant de s'affranchir des hallucinations observées avec la méthode RAG. A la suite de la notation, plusieurs graphiques ont été réalisés afin d'évaluer au mieux la fiabilité du programme.

Tableau 2 : Pourcentage général moyen représentant la part d'exactitude sur un compte-rendu généré par l'algorithme, en fonction de la version utilisée, associé à l'écart-type

	Version 1 - RAG	**Version 2 - GPT-3.5 Turbo**	**Version 2 - GPT-4 Turbo**	**Version 3 - GPT-4o**
Note moyenne	37,6/52	43,6/52	47,7/52	50,4/52
Ecart-type	∓ 5,9	∓2,9	∓2,6	∓1,7
Précision globale (%)	72,3 ∓ 11,3	83,8 ∓ 5,6	91,7 ∓ 5	96,9 ∓ 3,3

Dans un premier cadre global, on peut voir que chaque nouvelle version est plus précise et exacte que la précédente avec une moyenne plus élevée . De plus, dans le même temps que le LLM évolue, on remarque que l'écart-type diminue également. Ainsi, avec la méthode RAG, non seulement les comptes-rendus retournés n'étaient exacts que sur environ les trois quarts de leur contenu mais la variabilité aussi était élevée avec des comptes-rendus parfois fiables uniquement à 60%. A l'opposé, la dernière version de GPT (GPT-4o) mise à disposition par la société OpenAI®, a permis la rédaction de comptes-rendus rédigés avec les bonnes informations à hauteur de presque 97% de leur contenu et un écart-type inférieur à 2%.

Si on compare la version avec le RAG et la version faisant appel à GPT-3.5, le moins performant des "transformers", on peut voir que le taux moyen d'exactitude du "transformer" est de 10 points supérieure à celle du RAG. Cette différence s'explique par les modes de fonctionnement, dans un cas, avec la méthode retriever (RAG) et, dans l'autre cas, avec l'utilisation de "transformers" (GPT-3.4, 4 et 4o). En effet, comme expliqué plus tôt, les "transformers" permettent une meilleure compréhension de la sémantique là où le RAG fonctionne simplement par recherche de similarité de vecteur. Ce dernier cherche à relier les informations qui possèdent des vecteurs communs dans la transcription et dans la requête donnant ainsi lieu à des réponses parfois hors propos ou des bugs (figure 14). De plus, un autre problème rencontré avec le RAG, venant plutôt du LLM open source utilisé dans cette méthode, était la formulation de phrases dans un français peu soutenu et non professionnel, ou encore un manque de vocabulaire amenant le LLM à ne pas

comprendre certaines demandes. Nous pouvons notamment citer l'exemple de l'absence de distinction entre les antiparasitaires externes ou internes qui amenait souvent une réponse erronée sur la prise de renseignement concernant les vermifuges.

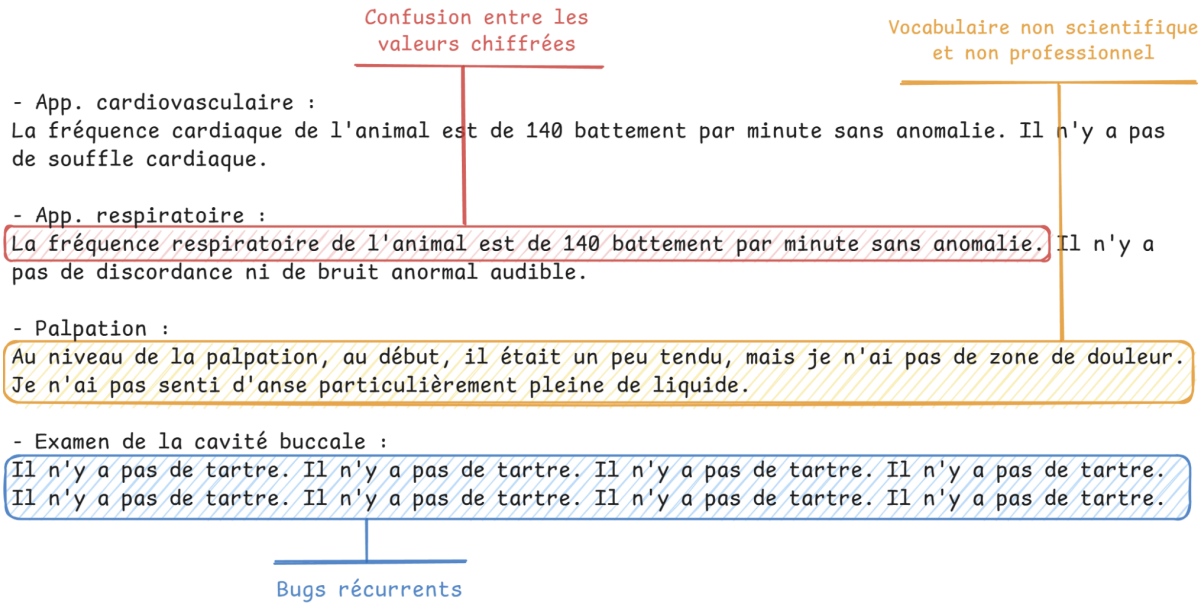


Figure 14 : Illustration d'un bug obtenu lors des itérations effectuées avec la méthode de génération augmentée par récupération

Lorsqu'on regarde le détail des moyennes des notations pour chaque requête (figure 15), on remarque que la méthode RAG est la moins performante sur presque l'entièreté des requêtes, celle concernant les API étant la moins exacte avec une moyenne à 0,94/2. La deuxième requête faisant baisser la moyenne générale est celle concernant l'anamnèse. Or, on remarque aussi que dans chacune des trois autres versions, l'anamnèse est toujours le facteur limitant du programme. En effet, cette requête a été la plus compliquée à formuler car elle collecte des informations très vastes. Chacune des autres requêtes était précise, avec une question fermée là où la requête concernant l'anamnèse amène une question ouverte avec, en plus de cela, une notion de temporalité, ce qui explique une justesse moins importante dans cette partie du compte-rendu.

Un autre résultat intéressant est la supériorité du modèle 3.5 Turbo par rapport au 4o et 4 Turbo concernant l'identification de la race de l'animal. Cela pourrait s'expliquer par un entraînement du LLM GPT 3.5 sur des données comprenant plus de mentions de races différentes d'animaux.

La figure 15 permet également de montrer que, bien que le GPT 3.5 montre des résultats plus satisfaisants que le système RAG, il se comporte tout de même comme ce dernier. En effet, les courbes sur le modèle RAG et sur le modèle GPT-3.5 Turbo sont presque superposables notamment sur toute la partie examen clinique, sur la partie concernant les caractéristiques physiologiques de l'animal ou encore son alimentation. Concernant l'anamnèse, on remarque que le transformer 3.5 était encore peu performant et avait des résultats semblables au système RAG.

Aussi, lorsqu'on observe les courbes représentant les modèles les plus récents, le 4 Turbo et le 4o, on remarque des courbes beaucoup plus lisse avec une exactitude, bien qu'imparfaite, toujours supérieure à 1,5/2 contrairement aux modèles précédents et ainsi des réponses toujours globalement proches de la réalité.

Enfin, le modèle 4o montre une supériorité ou, *a minima*, une égalité concernant l'exactitude de ses réponses à l'ensemble des requêtes par rapport aux autres modèles, excepté concernant la race de l'animal comme énoncé plus tôt. Ainsi, bien que l'anamnèse soit encore un facteur limitant sur le 4o, la moyenne de notation reste excellente avec une note global d'exactitude supérieure à 1,75/2.

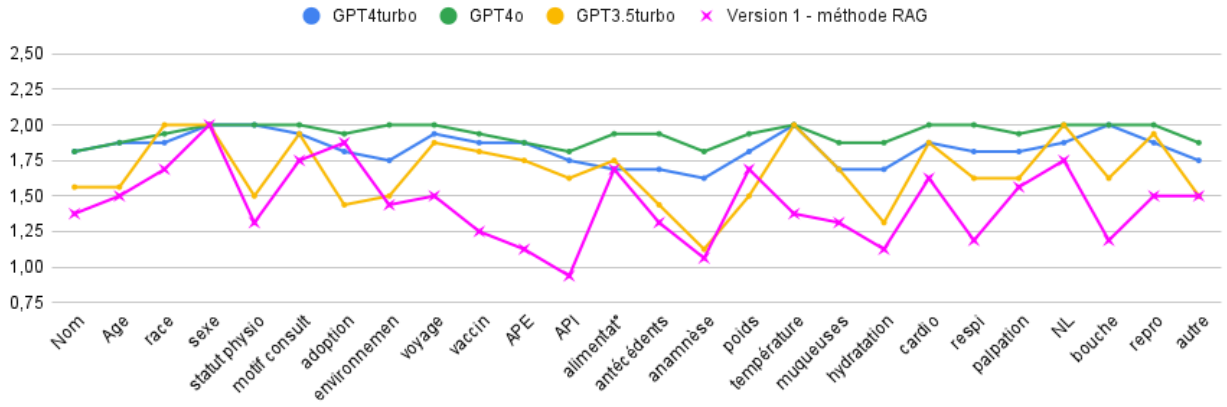


Figure 15 : Comparaison de l'exactitude des différentes versions de ReqVet en fonction de chaque requête

En combinant toutes les consultations, le modèle a répondu à un total de 416 requêtes et l'exactitude des réponses à l'ensemble de ces requêtes a été retranscrit sur la figure 16. Le but était d'observer si le fait que l'exactitude augmente avec l'évolution des modèles, était dû à une diminution du nombre de réponses fausses ou à des réponses globalement plus complètes sans variation du nombre de réponses fausses. On remarque ainsi que les réponses sont globalement plus

complètes avec une diminution de réponses partiellement exactes (note 1) montrant donc une capacité plus marquée du modèle à utiliser toutes les informations. De plus, le nombre de réponses fausses (note 0) diminue fortement avec les LLM les plus récents (figure 16). En effet, la différence entre la première version (RAG) et les suivantes avec les “transformers” est marquée avec une supériorité établie des modèles GPT illustrée par un nombre de réponses exactes (note 2) passant de deux tiers à plus de trois quarts. Enfin, concernant les trois différents “transformers” utilisés, on observe une augmentation nette du nombre de réponses exactes avec l’utilisation de modèles de plus en plus récents qui mènent à 95% de bonnes réponses avec le modèle GPT-4o.

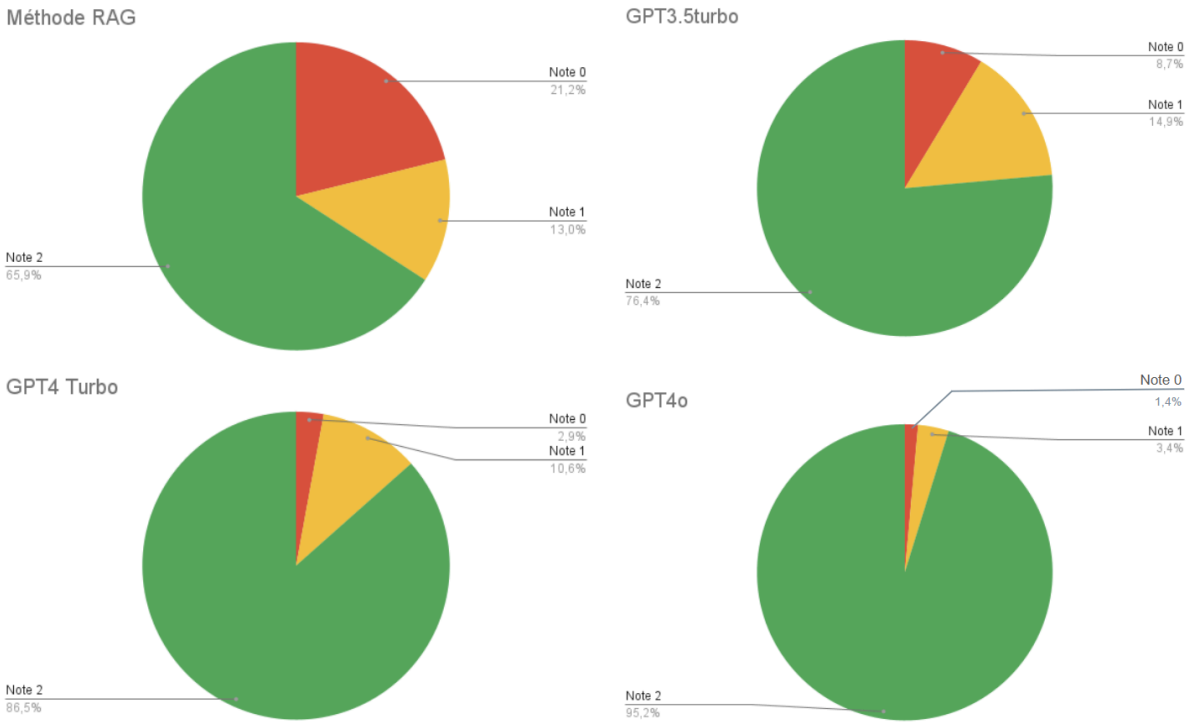


Figure 16 : Détail des notations sur les 416 requêtes posées (26 requêtes par consultation) avec chacune des versions de ReqVet utilisées

Il n’y pas de différence entre les GPT concernant le recueil des informations entre les parties anamnèse/commémoratifs et la partie examen clinique (figure 17). Or, on se serait attendu à observer plus de réponses exactes sur la partie examen clinique, dictée oralement et constituée uniquement de requêtes précises, là où les commémoratifs et l’anamnèse nécessitent d’apporter des informations plus complexes et issues d’un échange à base de questions/réponses avec

potentiellement des informations hors de propos liées aux échanges conversationnels de convenance. En définitive, dans notre étude, le modèle le plus récent est le seul à présenter des comptes-rendus présentant des performances égales sur l'examen clinique et l'échange verbal entre le propriétaire et le vétérinaire. Pour les autres, on remarque globalement que la moitié des comptes-rendus est plus juste sur la première partie et l'autre moitié, sur la deuxième.

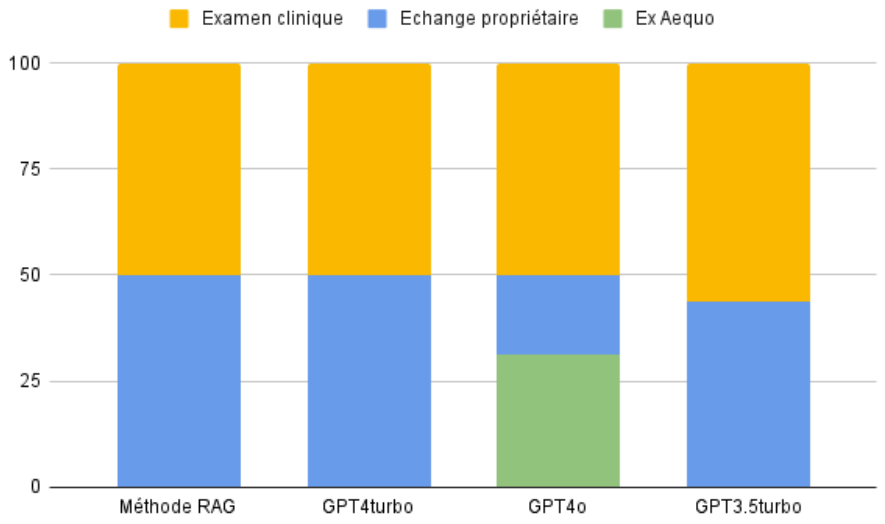


Figure 17 : Comparaison de l'exactitude des réponses selon une dichotomie du compte-rendu effectuée entre l'échange avec le propriétaire stricto sensu et la dictée de l'examen clinique

Ainsi, cette expérience a permis de constater une supériorité marquée de la dernière version de ReqVet avec des différences significatives d'un modèle sur l'autre montrant un compte-rendu globalement plus exact mais également une réponse individuelle à chaque requête améliorée. De plus, le nombre de réponses fausses étant fortement diminué, la fiabilité du programme est également établie avec une variabilité faible qui permet d'assurer un compte-rendu toujours fortement fidèle à la réalité peu importe le motif de consultation ou les conditions de celle-ci.

4. Discussion et perspectives

Bien que ReqVet, en l'état, offre déjà une bonne cohérence dans la génération de comptes-rendus, il est perfectible en plusieurs axes.

Pour commencer, il est important de souligner que, la rédaction du compte-rendu s'opérant à partir d'un audio, la partie portant sur l'examen clinique doit être dite à l'oral. Bien que ce dernier point soit une contrainte, le programme tel qu'il est, permet au vétérinaire de choisir, s'il le souhaite, de détailler l'examen clinique sur le mode d'une dictée vocale basique ou le dire directement à voix haute au moment où il le réalise.

Maintenant, l'IA n'étant pas réellement exhaustive ni fiable dans 100% des cas, couplé au fait que le compte-rendu ne se génère pas au fur et à mesure de la consultation mais seulement une fois l'échange terminé, chaque compte-rendu nécessite donc une relecture attentive. Il serait intéressant de chercher à réaliser la génération de compte-rendu de façon simultanée avec une transcription de l'audio en temps réel afin de pouvoir agir sur ce point.

De plus, le modèle de traitement du langage utilisé est, lui aussi, perfectible. En effet, il est possible de personnaliser le LLM, ici GPT-4o en l'occurrence, par ce qu'on appelle le "*fine-tuning*" : le fait d'entraîner le LLM sur des données choisies et ainsi, augmenter sa spécialisation et sa précision (IBM, 2024). Dans notre cas, il serait donc intéressant d'utiliser le "fine-tuning" afin de spécialiser le LLM dans différents domaines du monde vétérinaire ainsi qu'au vocabulaire propre à la profession. Ici, il s'agirait de donner au LLM de nombreux enregistrements de consultations associés à leur compte-rendu parfaitement rédigé ainsi de permettre au réseau de neurones une meilleure compréhension de notre requête et un ajustement de la réponse. De ce fait, non seulement cette méthode améliorerait la précision et la justesse du compte-rendu mais permettrait également d'adapter le vocabulaire avec des mots scientifiques adaptés et précis.

L'objectif sur le long terme de ReqVet, après avoir opéré un "fine-tuning" sur un grand échantillon de comptes-rendus, est de permettre au vétérinaire de sélectionner, avant chaque début de rendez-vous, le type de consultation qui va être réalisé (comportement, médecine, dermatologie, neurologie, imagerie…). De ce fait, il pourra avoir accès à un compte-rendu adapté à chaque cas avec un vocabulaire spécifique à la situation.

Actuellement, une fois le compte-rendu généré, une requête est soumise au LLM afin qu'il compare ce dernier à un petit dictionnaire de mots scientifiques. Ainsi, par similarité des vecteurs, l'IA opère un remplacement de l'expression commune utilisée par le terme scientifique adapté. Par exemple, "il boit beaucoup" devient "Polydipsie". Cependant, cette méthode est actuellement en cours d'amélioration et il semblerait que le manque de précision qui en ressort serait un autre argument en faveur de l'utilisation de "fine-tuning".

CONCLUSION

Force est de constater que l'intelligence artificielle intervient déjà dans de nombreux domaines du métier vétérinaire : aide au diagnostic, suivi des patients, interprétation d'examens complémentaires, alimentation… Cependant, la rédaction de compte-rendu, activité chronophage et pourtant essentielle à la bonne pratique de la profession, reste pour le moment absente de cette course à l'optimisation. Or, le compte-rendu est une tâche essentielle du métier de vétérinaire que ce soit sur un plan médical (suivi du patient) ou sur un plan juridique et comptable en y consignant les actes réalisés/proposés.

ReqVet a donc été créé afin d'optimiser la rédaction de comptes-rendus. En effet, cet outil rédige automatiquement des comptes-rendus vétérinaires sur la base de l'enregistrement de l'échange entre le professionnel de santé et le propriétaire et de l'énonciation de l'examen clinique à voix haute. Dans un premier temps, cet outil transcrit l'enregistrement vocal de la consultation en un fichier texte à l'aide de l'outil Whisper®. La seconde partie, la génération de compte-rendu *stricto sensu*, a été faite en deux temps : la première version utilisait un système de génération augmenté par récupération (méthode RAG) et permettait une exactitude moyenne de 72%. Afin d'améliorer la qualité des comptes-rendus générés, le système RAG a été remplacé par l'utilisation de grands modèles de langage (LLM) de type "transformers". Plus précisément, trois LMM ont été testé : GTP 3.5 turbo, GPT 4 turbo et GPT 4o jusqu'à obtenir, avec ce dernier, une exactitude moyenne de 97%.

Actuellement, ReqVet est encore en cours de perfectionnement, afin de pouvoir intégrer un vocabulaire scientifique adapté dans les anamnèses et les examens cliniques de façon constante et répétable. Le but étant de permettre son utilisation en clinique vétérinaire dans un avenir proche.

Pour finir, afin d'assurer une confidentialité maximale, tous les outils utilisés ont été étudiés pour répondre aux normes RGPD de sécurisation des données et ainsi être en adéquation avec le secret professionnel.

BIBLIOGRAPHIE

AMADIEU, Marie, 2023. *Analyse détaillée des performances d'une intelligence artificielle sur la détection radiographique des occlusions intestinales chez le chien et le chat*. Thèse d'exercice vétérinaire. Toulouse : Paul-Sabatier.

ANGEVERT, Luc, 2023. Avec l'IA générative, Nuance va transformer l'utilisation de Dragon en un copilote médical. *What's up Doc?* [en ligne]. Disponible à l'adresse : https://www.whatsupdoc-lemag.fr/grand-format/avec-lia-generative-nuance-va-transformer-lutilisation-de-dragon-en-un-copilote [Consulté le 29 mai 2024].

BOISSADY, Emilie, COMBLE, Alois, ZHU, Xiaojuan et HESPEL, Adrien-Maxence, 2020. Artificial intelligence evaluating primary thoracic lesions has an overall lower error rate compared to veterinarians or veterinarians in conjunction with the artificial intelligence. *Veterinary Radiology & Ultrasound*. Vol. 61. DOI 10.1111/vru.12912.

BURGER, Clarisse, SIGOT, Françoise, 2016. La transformation numérique des structures vétérinaires. *Le Point Vétérinaire.fr* [en ligne]. N° 1696. Disponible à l'adresse : https://www.lepointveterinaire.fr/publications/la-semaine-veterinaire/article/n-1696/la-transformation-numerique-des-structures-veterinaires.html [Consulté le 29 mai 2024].

COHERIS, 2022. Qu'est-ce que le Machine Learning ou apprentissage automatique ? *Intelligence artificielle & Data Analytics* [en ligne]. Disponible à l'adresse : https://ia-data-analytics.fr/machine-learning/ [Consulté le 28 mai 2024].

CORTADELLAS, Oscar, 2024. Taux créatinine chat : l'augmentation de la créatinine est-elle pertinente ? *Vets&Clinics* [en ligne]. Disponible à l'adresse : https://vetsandclinics.com/fr/taux-creatinine-chat-augmentation-creatinine-est-elle-pertinente [Consulté le 28 mai 2024].

DELL'ACQUA, Fabrizio, MCFOWLAND III, Edward, MOLLICK, Ethan R., LIFSHITZ-ASSAF, Hila, KELLOGG, Katherine, RAJENDRAN, Saran, KRAYER, Lisa, CANDELON, François et LAKHANI, Karim R., 2023. *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality* [en ligne]. SSRN Scholarly Paper. Rochester, NY. 4573321. [Consulté le 20 août 2024].

ELIAS, Greggory, 2024. Mistral 7B vs. LLama2 : Les 5 principales différences entre les principaux LLM open-source. *Skim AI* [en ligne]. Disponible à l'adresse : https://skimai.com/fr/mistral-7b-vs-llama2-les-5-principales-differences-entre-les-principaux-llm-open-source/ [Consulté le 11 juin 2024].

ESP, Earth Species Project, 2024. What We do. *ESP* [en ligne]. Disponible à l'adresse : https://www.earthspecies.org/what-we-do/technology [Consulté le 27 mai 2024].

FÉVRIER, Lucie, 2017. *La comptabilisation du temps de travail du vétérinaire libéral : étude du « système de points »*. Thèse d'exercice vétérinaire. Toulouse : Paul-Sabatier.

FINNEGAN, Joanne, 2020. For each patient visit, physicians spend about 16 minutes on EHRs, study finds. *Fierce Healthcare* [en ligne]. Disponible à l'adresse : https://www.fiercehealthcare.com/practices/for-each-patient-visit-physicians-spend-about-16-minutes-ehrs-study-finds [Consulté le 30 mai 2024].

GALLAND, Joris, 2020. La médecine des 4P... et plus. *Esanum* [en ligne]. Disponible à l'adresse : https://www.esanum.fr/today/posts/la-medecine-des-4p-et-plus [Consulté le 28 mai 2024].

GALLOWAY, Lucy, 2023. Outils de dictée médicale : 5 avantages pour les cabinets médicaux. *Pabau* [en ligne]. Disponible à l'adresse : https://pabau.com/fr/blog/5-avantages-de-lutilisation-doutils-de-dictee-medicale-dans-votre-cabinet-prive/ [Consulté le 29 mai 2024].

GAO, Yunfan, XIONG, Yun, GAO, Xinyu, JIA, Kangxiang, PAN, Jinliu, BI, Yuxi, DAI, Yi, SUN, Jiawei, WANG, Meng et WANG, Haofen, 2024. *Retrieval-Augmented Generation for Large Language Models: A Survey* [en ligne]. Disponible à l'adresse : http://arxiv.org/abs/2312.10997 [Consulté le 24 juillet 2024].

GITHUB, 2024. Model-card of whisper training by OpenAI. *github.com* [en ligne]. Disponible à l'adresse : https://github.com/openai/whisper/blob/main/model-card.md [Consulté le 10 juin 2024].

HERVEY-CHUPIN, Diane, 2022. Données et secret professionnel. *Bulletin de l'Académie Vétérinaire de France*. Vol. 175, pp. 73-83. DOI 10.3406/bavf.2022.70993.

IBM, 2024. Qu'est-ce qu'un grand modèle de langage (LLM) ? *IBM* [en ligne]. Disponible à l'adresse : https://www.ibm.com/fr-fr/topics/large-language-models [Consulté le 6 juin 2024].

IDEXX, 2024. L'intelligence artificielle appliquée à la médecine vétérinaire mène à une efficacité et une précision accrues. [en ligne]. Disponible à l'adresse : https://ca.idexx.com/fr-ca/veterinary/analyzers/artificial-intelligence-veterinary-medicine-leads-to-efficiency/ [Consulté le 27 mai 2024].

JEANVIET, 2023. *Whisper d'Open AI : Comment transformer de l'audio en texte gratuitement ?* [en ligne]. Disponible à l'adresse : https://jeanviet.fr/whisper/ [Consulté le 10 juin 2024].

KIEFFER, Jean-Pierre, 2014. Le secret professionnel en médecine vétérinaire. *Le Point Vétérinaire.fr* [en ligne]. N° 1590. Disponible à l'adresse : https://www.lepointveterinaire.fr/publications/la-semaine-veterinaire/article-asv/n-1590/le-secret-professionnel-en-medecine-veterinaire.html [Consulté le 6 août 2024].

LAMOLY, Amélie, 2020. *L'application de l'intelligence artificielle au service de la nutrition individualisée*. Thèse d'exercice vétérinaire. Toulouse : Paul-Sabatier.

MASSON, Laurent, 2020. Examen complémentaires lors de maladie rénale chronique chez le chat. *Le Point Vétérinaire.fr* [en ligne]. N° 1843. Disponible à l'adresse : https://www.lepointveterinaire.fr/publications/la-semaine-veterinaire/article/n-1843/examens-complementaires-lors-de-maladie-renale-chronique-chez-le-chat.html [Consulté le 28 mai 2024].

MÉDARD, François-Henri, 2007. Quatre logiciels de gestion du cabinet vétérinaire sortent du lot, comme il y a deux ans. *Le Point Vétérinaire.fr* [en ligne]. N° 1263. Disponible à l'adresse : https://www.lepointveterinaire.fr/publications/la-semaine-veterinaire/article/n-1263/quatre-logiciels-de-gestion-du-cabinet-veterinaire-sortent-du-lot-comme-il-y-a-deux-ans.html [Consulté le 29 mai 2024].

MÉDARD, François-Henri, 2013. Logiciels de gestion de cabinet : nouveaux services, nouveaux acteurs. *Le Point Vétérinaire.fr* [en ligne]. N° 1534. Disponible à l'adresse : https://www.lepointveterinaire.fr/publications/la-semaine-veterinaire/archives/n-1534/logiciels-de-gestion-de-cabinet-nouveaux-services-nouveaux-acteurs.html [Consulté le 30 mai 2024].

MISTRALAI, 2023. *Mistral 7B*. [en ligne]. Disponible à l'adresse : https://mistral.ai/fr/news/announcing-mistral-7b/ [Consulté le 11 juin 2024].

MOTTAY, Nicolas Chauchart Du, 2023. *Utilisation de l'Intelligence Artificielle dans la gestion des maladies chroniques des animaux de compagnie: opportunités et freins à l'adoption*. Thèse d'exercice vétérinaire. Créteil : Faculté de médecine UPEC.

NUANCE COMMUNICATIONS, 2024. *Dragon Medical One | Documentation Clinique AI Assistant.* [en ligne]. 2024. Disponible à l'adresse : https://www.nuance.com/fr-fr/healthcare/provider-solutions/speech-recognition/dragon-medical-one.html [Consulté le 29 mai 2024].

OPENAI, 2024. *OpenAI Security Portal* [en ligne]. Disponible à l'adresse : https://trust.openai.com/ [Consulté le 19 août 2024].

OPENAI, 2024. *OpenAI Platform.* [en ligne]. Disponible à l'adresse : https://platform.openai.com [Consulté le 24 juillet 2024].

PERRIN, Roxanne, 2019. *Emergence de l'intelligence artificielle et utilisation des technologies Big Data en médecine vétérinaire : importance de la sensibilisation des futurs vétérinaires*. Thèse d'exercice vétérinaire. Créteil : Faculté de médecine UPEC.

PONARD, Audrey, 2023. *Évaluation de l'efficacité et perfectionnement d'un outil d'aide au diagnostic et de classification en pathologie des ruminants: application aux affections mammaires, circulatoires, cutanées et plus*. Thèse d'exercice vétérinaire. Lyon : Université Claude Bernard Lyon 1 (Médecine – Pharmacie).

PROMPT ENGINEERING GUIDE, 2024a. *Génération Augmentée par Récupération (RAG).* [en ligne]. Disponible à l'adresse : https://www.promptingguide.ai/fr/techniques/rag [Consulté le 24 juillet 2024].

PROMPT ENGINEERING GUIDE, 2024b. *Prompt Chaining.* [en ligne]. Disponible à l'adresse : https://www.promptingguide.ai/fr/techniques/prompt_chaining [Consulté le 24 juillet 2024].

RADFORD, Alec, KIM, Jong Wook, XU, Tao, BROCKMAN, Greg, MCLEAVEY, Christine et SUTSKEVER, Ilya, 2022. Robust Speech Recognition via Large-Scale Weak Supervision. *International Conference on Machine Learning*, juillet 2023 à Honolulu. JMLR.org, Article No. 1182, Pages 28492 - 28518

RUNPOD.IO, 2024. *Compliance and Security at RunPod.* [en ligne]. Disponible à l'adresse : https://www.runpod.io/compliance#compliance-overview [Consulté le 8 août 2024].

RUTZ, Christian, BRONSTEIN, Michael, RASKIN, Aza, VERNES, Sonja C et BLASI, Damián E, 2023. Using machine learning to decode animal communication. *Science.* Vol. 381, n° 6654, pp. 152.

SCHILTZ, Louis-Clément, 2023. *Clé de l'IA Chatgpt: Fonctionnement et Impact des Grands Modèles de Langage.* [en ligne]. Disponible à l'adresse : https://www.webotit.ai/cle-de-lia-chatgpt-fonctionnement-et-impact-des-grands-modeles-de-langage [Consulté le 12 juin 2024].

TALAMADUPULA, Kartik, 2024. A Guide to Quantization in LLMs. *Symbl.ai* [en ligne]. Disponible à l'adresse : https://symbl.ai/developers/blog/a-guide-to-quantization-in-llms/ [Consulté le 12 juin 2024].

TALBI, Ilyes, 2020. Comprendre les réseaux de neurones. *La revue IA* [en ligne]. Disponible à l'adresse : https://larevueia.fr/comprendre-les-reseaux-de-neurones/ [Consulté le 29 mai 2024].

TALBI, Ilyes, 2023. Introduction aux réseaux de neurones Transformers. *La revue IA* [en ligne]. Disponible à l'adresse : https://larevueia.fr/introduction-aux-reseaux-de-neurones-transformers/ [Consulté le 12 juin 2024].

TALKATOO, 2024. Talkatoo Dictation Software | Save Time, Get More Done. *Talkatoo* [en ligne]. Disponible à l'adresse : https://talkatoo.com/ [Consulté le 29 mai 2024].

VETOCOM, 2024. Mobilité - mVet par Vétocom - L'application connectée à votre logiciel. *vetocom* [en ligne]. Disponible à l'adresse : https://www.vetocom.fr/fonctionnalites/vetocom-autres-fonctionnalites/mobilite-mvet/ [Consulté le 29 mai 2024].

VIKINGGENETICS, 2020. Artificial intelligence makes cows more feed efficient and climate-friendly. *VikingGenetics* [en ligne]. Disponible à l'adresse : https://www.vikinggenetics.com/press-releases/pr-artificial-intelligence [Consulté le 28 mai 2024].

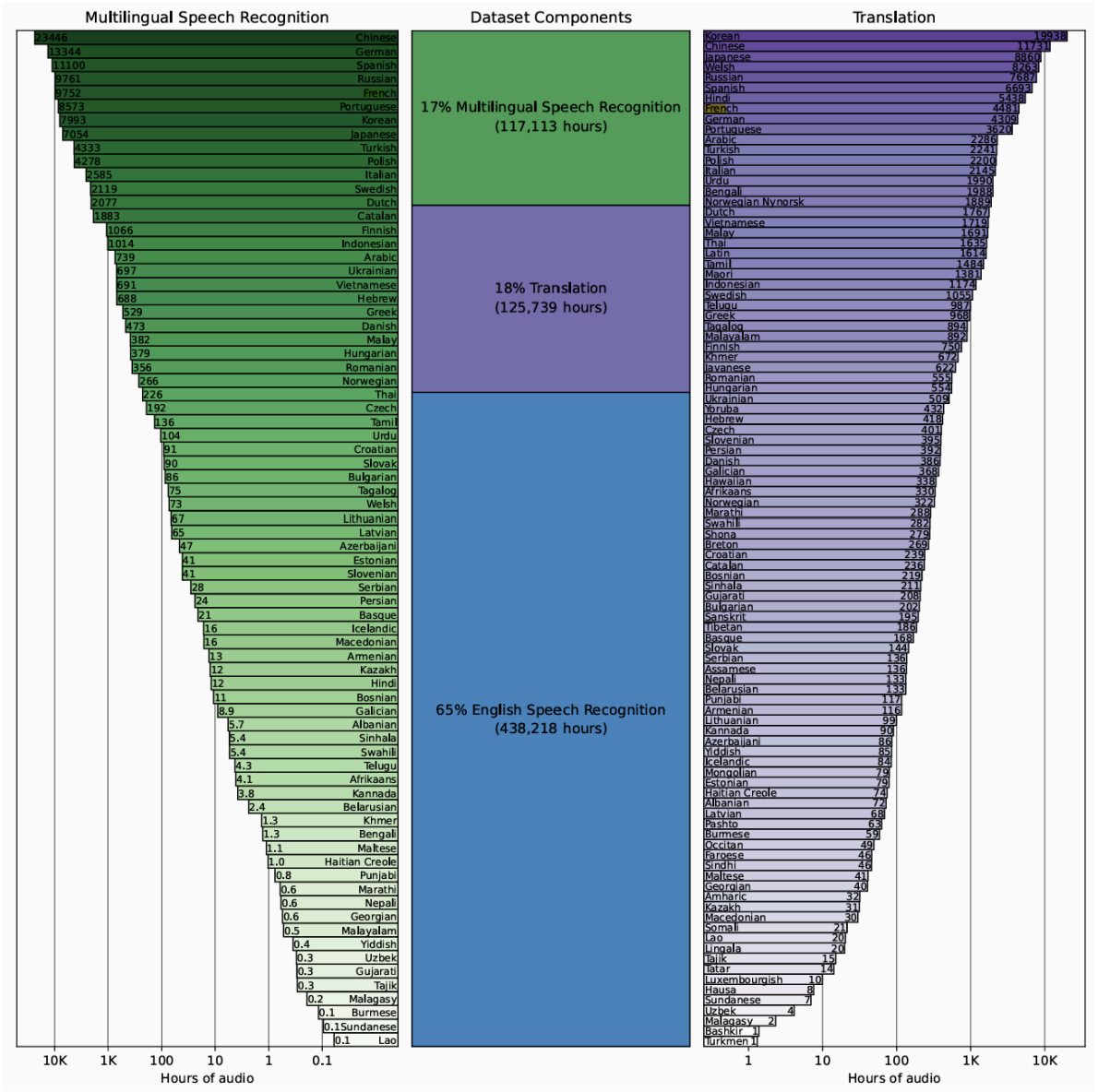
ZOETIS FRANCE, 2024. *Vetscan Imagyst* [en ligne]. 2024. Disponible à l'adresse : https://www2.zoetis.fr/vetscan-imagyst/ [Consulté le 27 mai 2024].

ZOETIS US, 2024. AI Dermatology. *Vetscan Imagyst* [en ligne]. 2024. Disponible à l'adresse : https://www.VetscanImagyst.com/ai-dermatology [Consulté le 26 septembre 2024].

ANNEXES

ANNEXE 1 : **Statistique des données d'entraînement de Whisper®** - [whisper.pdf (openai.com)]

E. Training Dataset Statistics



<u>ANNEXE 2</u> : **Modèle général de compte-rendu vétérinaire basé sur les modèles SIRIUS®.**

COMPTE-RENDU VÉTÉRINAIRE

<u>Commémoratifs :</u>

- Nom de l'animal :
- Âge :
- Race :
- Sexe :
- Statut physiologique :

- Motif de la consultation

- Adoption :
- Environnement :
- Voyages :
- Statut vaccinal :
- Antiparasitaire externe :
- Antiparasitaire interne :
- Alimentation :
- Antécédents médicaux :

<u>Anamnèse :</u>

<u>Examen Clinique :</u>

- Poids (et score corporel si indiqué) :
- Température :
- Muqueuses :
- Hydratation :
- Appareil cardiaque :
- Appareil respiratoire :
- Palpation et noeuds lymphatiques :
- Cavité buccale :
- Appareil reproducteur :
- Autre :

<u>ANNEXE 3</u> : Exemple d'un compte-rendu retourné par ReqVet après une consultation

I. Commémoratif :

- **Description de l'animal :**

  * *Nom* : Lilou

  * *Age* **:** 8 ans

  * *Race* : Burmese

  * *Sexe* : Femelle

  * Statut physiologique : Non précisé

- Raison de la visite : Le propriétaire apporte son animal car Lilou présente des difficultés à uriner, avec seulement quelques gouttes d'urine visibles, et elle se rend fréquemment à la litière sans succès. Le propriétaire suspecte une récidive de cystite.

- Adoption / Acquisition : Non précisé

- **Cadre de vie** : Lilou vit en appartement et ne sort pas. Elle est le seul animal mentionné dans le foyer.

- **Voyages** : Non précisé

- Statut vaccinal : Non précisé

- Traitement antiparasitaires :

Antiparasitaire externe (ou APE) : Frontline Combo, appliqué tous les mois.

Antiparasitaire interne (ou API) : Comprimés vermifuges administrés tous les deux mois, nom non précisé.

- Alimentation : Lilou est nourrie avec des croquettes Royal Canin Urinary depuis trois jours. Elle a également reçu des croquettes Optimum et des sachets Optimum le matin.

- Antécédents : Lilou a eu un épisode de cystite début septembre, traité avec des antibiotiques, du méloxicam, du spasfon et une alimentation urinaire. Elle a également eu un problème de paralysie des membres postérieurs dans le passé, traité par physiothérapie, mais cela s'est résolu.

II. Anamnèse :

Lilou présente des symptômes de dysurie depuis 3 à 4 jours, avec une absence d'urine visible depuis hier. Elle se rend fréquemment à la litière sans succès. Le propriétaire a commencé à lui donner des croquettes Royal Canin Urinary depuis trois jours. Lilou ne semble pas miauler de douleur, mais elle se met en position pour uriner sans succès. Elle mange et boit normalement. Les selles ont été moins fréquentes et en petite quantité récemment.

III. Examen Clinique :

Général :

* *Poids (et score corporel, si notifié)* : Score corporel de 7 sur 9

* *T°C corporelle* : Non précisée

* *Etat des muqueuses* : Très pâles et humides

* *Etat d'hydratation* : Non déshydratée

- App. Cardiovasculaire : Fréquence cardiaque à 180 BPM, pas d'anomalie à l'auscultation, TRC non évaluable.

- App. Respiratoire : Fréquence respiratoire à 36 mpm, sans anomalie à l'auscultation.

- Palpation : Non précisée

Noeuds Lymphatiques : Pas d'adénomégalie périphérique

- Examen de la cavité buccale : Non précisé

- App. reproducteur : Non précisé

- Autre : Lilou est trop stressée pour un examen complet.

<u>ANNEXE 4</u> : Extrait, traduit en français, d'un échange par mail réalisé avec la plateforme Runpod.io® renseignant sur leur politique de sécurisation des données.

- **Certifications de conformité**: RunPod cherche activement à obtenir la certification SOC 2 Type 1 à l'échelle de la plate-forme en 2024, avec des plans pour la conformité SOC 2 Type 2, ISO, HIPAA et GDPR d'ici la fin de 2025. Certains de nos centres de données disposent déjà de certifications étendues, mais ce n'est pas encore le cas de la plateforme cloud RunPod dans son ensemble.
- **Centres de données conformes au RGPD** : Nos centres de données européens répondent aux normes GDPR. La conformité peut être assurée en utilisant des serveurs exclusivement en Europe - cette option est disponible à la fois sur nos plateformes Secure Cloud et Serverless. Assurez-vous simplement de filtrer ces emplacements lorsque vous louez des ressources sur RunPod.

<u>ANNEXE 5</u> : Formulaire de consentement présenté aux propriétaires afin de permettre l'enregistrement des consultations

CENTRES DE RECHERCHE ET RESSOURCES DOCUMENTAIRES

DÉCLARATION DE CONSENTEMENT ÉCLAIRÉ CONSULTATION SOUS DISPOSITIF D'ENREGISTREMENT AUDIO

Je, soussigné(e), Madame, Monsieur : ……

Demeurant : …….

Téléphone : ……..

☐ Déclare par la présente accepter, librement, de façon éclairée et univoque, de participer à une consultation soumise à un enregistrement audio impliquant que des données à caractère personnel me concernant fassent l'objet d'un traitement et avoir bien pris connaissance du projet dans lesquelles ces données seront utilisées.

☐ Je comprends que les données directes permettant de m'identifier ne seront utilisées que dans le cadre des activités de recherche de l'École Vétérinaire de Toulouse, qu'elles seront soumises à confidentialité et que ces données ne pourront être utilisées que dans le contexte d'un processus d'élaboration d'une thèse vétérinaire dénommée « Utilisation de l'intelligence artificielle pour la rédaction de compte-rendu vétérinaire automatisé à partir d'un fichier audio » ; thèse visant, à partir de l'enregistrement audio d'un échange entre propriétaire et vétérinaire, à permettre la rédaction de compte-rendu vétérinaire rapide et le plus exhaustif possible.

☐ J'ai connaissance de mon droit de retirer mon consentement à tout moment. Il est aussi simple de retirer que de donner son consentement. Le retrait du consentement ne compromet pas la licéité du traitement fondé sur le consentement effectué avant ce retrait.

Fait à Signature :

le

Conformité au Règlement Général sur la Protection des Données.

Vous avez accepté de participer à un entretien filmé impliquant de répondre à un certain nombre de questions en présence d'un dispositif d'enregistrement audio. Cette consultation enregistrée constitue un traitement de vos données personnelles. Les données collectées seront traitées par les chercheurs en charge de la réalisation de la thèse vétérinaire intitulée «Utilisation de l'intelligence artificielle pour la rédaction de compte-rendu vétérinaire automatisé à partir d'un fichier audio » et utilisées uniquement à cette fin.

Conformément au Règlement Général sur la Protection des Données, vos données seront traitées pour la durée nécessaire à la réalisation de l'objectif poursuivi par la collecte, soit, une durée de 10 mois *(dix mois)* dans la présente étude. Au-delà de ce délai, elles seront supprimées.

Pour toute question relative au traitement de vos données personnelles ou pour exercer vos droits, vous pouvez écrire au Délégué à la protection des données personnelles de l'établissement à l'adresse suivante : dpo@envt.fr

Sandra FOUREL

ESSAI D'UTILISATION DE L'INTELLIGENCE ARTIFICIELLE POUR RÉDIGER DES COMPTES-RENDUS EN MÉDECINE VÉTÉRINAIRE

Cette thèse détaille la création d'un outil utilisant l'intelligence artificielle (IA) pour automatiser la rédaction des comptes-rendus en médecine vétérinaire. Elle commence par un état des lieux détaillé des diverses applications de l'IA dans le domaine vétérinaire. Ensuite, elle présente l'outil ReqVet, conçu pour générer des comptes-rendus vétérinaires à partir d'enregistrements vocaux de consultations. L'étude réalisée explique la démarche et les choix effectués au cours de la mise en place de ReqVet puis compare différents modèles de traitement du langage naturel, en évaluant leur précision, leur efficacité et leur capacité à traiter des données spécifiques au domaine vétérinaire. Le tout est développé dans le respect des normes RGPD pour garantir la confidentialité des données. Ce projet a pour objectif de réduire la charge administrative des vétérinaires, tout en améliorant la qualité des écrits et le suivi des patients.

Mots-clés : INTELLIGENCE ARTIFICIELLE - COMPTES-RENDUS - ENREGISTREMENT VOCAL - CONSULTATION - VÉTÉRINAIRE

ATTEMPT TO USE ARTIFICIAL INTELLIGENCE FOR WRITING REPORTS IN VETERINARY MEDICINE

This thesis details the creation of a tool using artificial intelligence (AI) to automate the writing of reports in veterinary medicine. It begins with a comprehensive overview of various applications of AI in the veterinary field. Next, it presents the tool ReqVet, designed to generate veterinary reports from voice recordings of consultations. The study explains the approach and choices made during the implementation of ReqVet, and then compares different natural language processing models, evaluating their accuracy, efficiency, and ability to handle domain-specific data. The entire project is developed in compliance with GDPR regulations to ensure data confidentiality. The objective of this project is to reduce the administrative burden on veterinarians while improving the quality of documentation and patient follow-up.

Keywords: ARTIFICIAL INTELLIGENCE - REPORTS - VOICE RECORDING - CONSULTATION - VETERINARY