

# Projet tableau de bord Groupe 8

## 2020/2021

Arouk Amine  
Azougagh Fouad  
Benhamouche Sofiane  
Dahmi Issam  
Hadda Hassan

Mme Wahiba Bahsoun  
M Riad Mokadem

## Table des matières

<b>Objectif du document .....</b>	<b>3</b>
1 - Objet .....	
2 - Présentation générale du document .....	3
2.1 - But .....	3
2.2 - A qui il est destiné .....	3
2.3 - Problématique étudiée .....	3
<b>Documents de références et terminologie .....</b>	<b>4</b>
1 - Documents de références .....	4
2 - Documents applicables .....	4
3 - Terminologie.....	5
<b>Organisation de développement.....</b>	<b>6</b>
1 - Fonctionnement du groupe.....	6
2 – Ressources humaines .....	7
3 – Répartition des tâches.....	8
<b>Démarche et développement.....</b>	<b>9</b>
1 – Recherche et collecte .....	9
2 – Préparation.....	11
3 – Valorisation des données .....	12
4 – Diffusion des connaissances .....	13
4.1 – Le nuage des mots du corpus .....	14
4.2 – Les pays de publication par maladie .....	15
4.3 – Les méthodes d’intelligence artificielle utilisées .....	16
4.4 – Evolution du nombre de parution d’articles.....	17
4.5 – Fréquence des mots clés .....	18
4.6 – Evolution du nombre de parution d’articles.....	19
4.7 – Graphique regroupant l’ensemble des paramètres étudiés.....	19
<b>Gestion de configuration .....</b>	<b>21</b>
1 – Ressources matérielles .....	21
2 – Les différents versions exploitées .....	25
<b>Assurance qualité.....</b>	<b>25</b>
1 – Revues .....	25
<b>Conclusion et bilan du projet .....</b>	<b>27</b>
1 – Conclusion .....	27
2 – Bilan.....	28
<b>ANNEXE.....</b>	<b>30</b>

# Objectif du document

## 1. Objet

Ce projet “tableau de bord” a été mené par cinq étudiants en première année de Master en Statistiques et Informatique Décisionnelle de l’Université Paul Sabatier à Toulouse. Encadré par deux professeurs, l’objectif a été de répondre à une problématique au choix par le biais de différents graphiques. Pour ce faire, les interprétations ont portées sur des analyses textuelles d’articles disponibles sur une bibliothèque scientifique en ligne.

## 2. Présentation générale du document

### 2.1 But

Au fur et à mesure de l’avancement du projet, ce rapport a été conçu dans le but d’expliquer la démarche entreprise par l’équipe en présentant la façon de procéder, les difficultés rencontrées, ainsi que les résultats.

### 2.2 A qui il est destiné

Par conséquent, ce document est destiné aux clients, qui sont dans ce cas nos tuteurs de projet Madame Wahiba Bahsoun et Monsieur Riad Mokadem. Le but étant qu’ils évaluent notre gestion de projet ainsi que les résultats obtenus.

### 2.3 Problématique étudiée

L'intelligence artificielle (IA) est un domaine de recherche en pleine expansion et promis à un grand avenir. Ses applications, qui concernent toutes les activités humaines, permettent notamment d'améliorer la qualité des soins. L'IA est en effet au cœur de la médecine du futur, avec les opérations assistées, le suivi des patients à distance, les prothèses intelligentes, les traitements personnalisés grâce au recoupement d'un nombre croissant de données. Par conséquent, nous avons choisis de traiter la problématique suivante :

Quelle est la part contribution de l'IA dans les maladies incurables

## **Documents de référence et terminologie**

### **1. Documents de références**

Durant tout le déroulement du projet il nous a fallu mettre en application beaucoup de compétences acquises durant ces années de formation que ce soit en SID ou nos premières années de licences au travers des cours, TPs et projets que l'on a pu effectuer.

Ce projet faisant parti des plus longs que l'on ait pu faire dans cette formation, il n'a pas fallu négliger l'aspect organisation du groupe et des méthodes de travail que l'on a pu découvrir durant les cours de processus de développement et de génie logiciel.

Pour la phase d'extraction et traitement de texte nous avons pu mettre en pratique les cours d'extraction d'informations, de Data et Text Mining ainsi que les TPs correspondants mais également les cours d'Informatique auxquels vous avons pu assister avant cette formation.

Pour la partie stockage, implémentation et maintenance de la base de données, les cours et TPs de base de données ont montrés leur utilité et nous ont permis de résoudre certains soucis plus rapidement.

Des sites comme Stackoverflow, Openclassrooms, Developpez, W3school nous étaient alors déjà connu et nous ont beaucoup apporté pour la partie codage et récolte de données.

### **2. Documents applicables**

Il a fallu s'assurer d'être en accord tout au long du projet avec les besoins du client. Pour ce faire, à chacune des grandes étapes du projet nous nous sommes basés sur le Cahier des Charges qui a ainsi orienté nos décisions concernant la conception, l'affichage du tableau

de bord final ainsi que pour le fonctionnement du groupe et des différentes traces écrites à conserver pour attester de la qualité du travail fourni.

Le tableau de bord devant présenter un ensemble de graphiques dans sa finalité, il nous a fallu récupérer un grand nombre d'articles afin de mener des analyses pertinentes. Ainsi l'ensemble des informations autour de ces articles ont dû être stockés dans une base de données dont nous avons dû penser un modèle entité-association et relationnel mettant en jeu le respect des normes NF.

### **3. Terminologie**

MCD : Modèle Conceptuel des Données

Scrapping : Récupérer le code source d'une page

A.I : Artificial Intelligence

Dashboard : Tableau de Bord

Dataframe : Objet python pour conserver des données sous forme de table

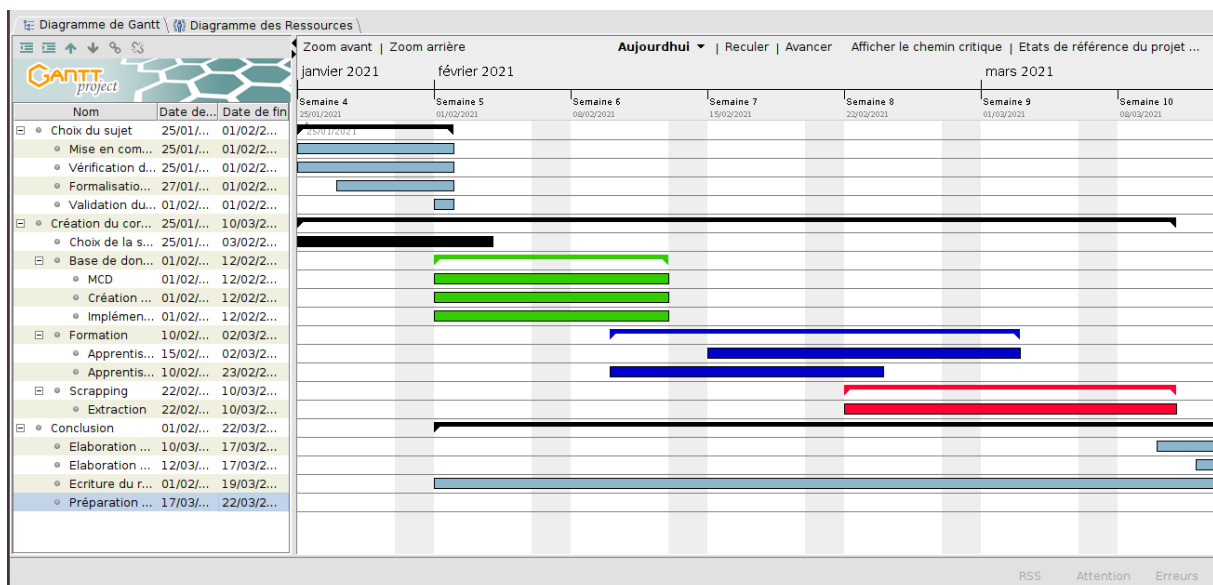
# Organisation de développement

## 1. Fonctionnement du groupe

Plusieurs contraintes ont vu le jour dès le début du projet, tout d'abord le fait que tout se déroule en distanciel et qui pourrait ainsi perturber la communication du groupe mais aussi le fait que le projet s'étale sur plusieurs mois et nécessite d'être traité sur plusieurs aspects. Nous avons alors opté pour un diagramme de GANTT pour structurer les missions qui allaient nous être demandé et ainsi garder un œil sur l'évolution du projet dans sa globalité.

On a essayé d'estimer une durée pour chaque point et éviter de déborder et se retarder pour la suite. Des tâches qui nous avaient échappées au début ont été rajoutées tout au long du projet.

Bien que la capture du diagramme (cf schema suivant) soit la version finale du Gantt, les premières versions faisaient aussi apparaître un grand nombre de tâches à effectuer, nous avons donc conclu qu'il serait intéressant de programmer des réunions de mise au point chaque semaine pour mettre en commun ce que chacun avait pu produire.



On avait ainsi deux réunions par semaine :

- Une réunion tous les samedis avec un point sur ce qui a été fait et ce qui doit être fait pour le cours suivant.
- Une réunion pendant les cours avec les enseignants qui nous corrigent ce qui ne va pas et nous réorientent.

Cela a plutôt bien fonctionné étant donné que l'on était très souvent dans les temps et qu'on a pu fournir un Dashboard en livrable et non seulement des graphiques.

Les réunions étaient programmées sur Discord pour faciliter le partage de plusieurs écrans en simultanée, on pouvait ainsi aussi mettre nos caméras pour essayer d'effacer le côté réunion virtuelle qui peut devenir assez fatigant sur le long terme.

Pour les sessions de travail personnel, on a mis à disposition plusieurs fichiers moins formels où l'on pouvait inscrire toutes nos idées (notamment les idées de graphiques, les façons d'extraire et nettoyer les données) et noter les avancées de chacun. Pour les documents plus formels (les différentes versions de code, du MCD et du rapport), un espace GitHub a été créée avec une section par membre pour déposer ses codes et une section générale avec les dernières versions de chaque document.

## **2. Ressources humaines**

En dehors de l'aspect programmation du projet, chacun d'entre nous s'est vu assigné une fonction. Le projet a alors été réalisé par cinq étudiants, ayant des rôles différents :

- Chef de projet :  
Ce rôle a été attribué à Fouad Azougagh
- Responsable assurance et contrôle qualité  
Ce rôle a été attribué à Amine Arouk
- Responsables de configuration  
Ce rôle a été attribué à Sofiane Benhamouche et Hassan Hadda
- Responsable de rédaction  
Ce rôle a été attribué à Issam Dahmi

### **3. Répartition des tâches**

Pour la phase préliminaire de recherches de la problématique et de l'équation de recherche correspondante ainsi que les phase de nettoyage et de traitement des données, nous avons estimé que chacun d'entre nous devait y prendre part soit car elles avaient une influence sur l'objectif du projet soit car elles nécessiteraient du travail tout le long du projet. Pour ces phases là on retrouve donc chacun des membres du groupe qui a pu participer.

Pour la création et maintenance de la base de données, l'extraction et l'affichage des données nous avons estimé qu'elles dureraient au maximum 2 semaines et nous les avons donc répartis respectivement entre Sofiane, Amine & Fouad, Hassan & Issam. L'objectif n'était pas que seulement quelques personnes s'y intéressent mais que la grosse partie du codage correspondante ne soit effectuée que par une partie d'entre nous, le reste du groupe étant tout de même présent pour aider au nettoyage et au traitement des données.

L'objectif derrière cela était de distribuer équitablement les missions en fonction de si on les avait estimées avec une durée limitée dans le temps ou celles qui allaient devoir être maintenues tout le long. Ces tâches se font bien évidemment rajoutées en parallèle des rôles assignés auparavant.

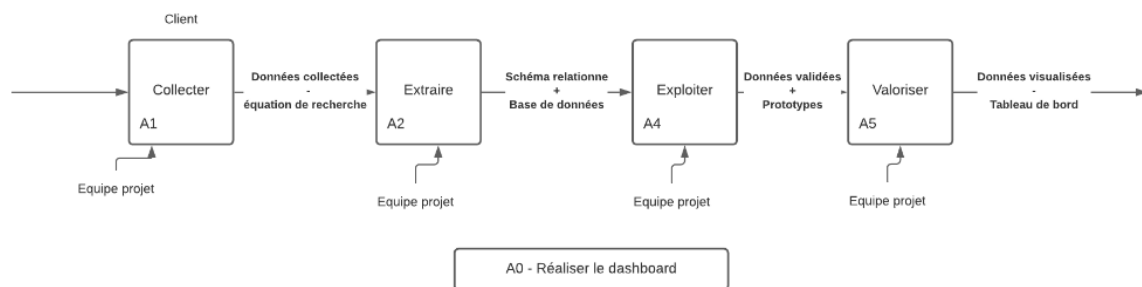
Un récapitulatif de la distribution de ces tâches :

	Les différentes phases du projet				
	Création & maintenance de la base de données	Extraction des données	Nettoyage des données	Traitement des données	Affichage des données
Arouk Amine		X	X	X	
Azougagh Fouad		X	X	X	
Benhamouche Sofiane	X		X	X	X
Dahmi Issam			X	X	
Hadda Hassan			X	X	X

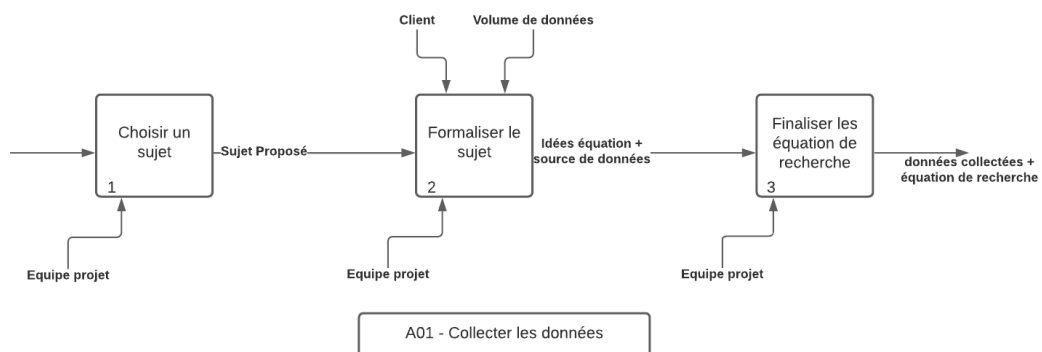


# Démarche et développement

Nous avons vu précédemment les différentes phases du processus de travail, l'idée est ici d'expliquer plus en détail ce qui a été mené durant chacune des étapes résumées dans le SADT suivant :



## 1. Recherche et collecte des données



Une fois notre sujet choisi et validé, la première étape a été de trouver une source de données pertinente pour répondre à notre problématique. Après plusieurs recherches, notre choix s'est fixé sur le site PubMed, une bibliothèque en ligne présentant un nombre d'articles intéressant dans le domaine scientifique.

Afin de se faire une idée sur les graphiques à mettre en avant dans le tableau de bord, nous avons étudié le contenu de plusieurs articles pour cibler les différents axes d'analyse

possibles. Dès lors, nous avons pu établir une première version de notre MCD qui sera mise à jour au fur et à mesure de nos nouvelles idées.

L'étape suivante avec la mise en place de notre équation de recherche a été très importante pour la suite, il a fallu peu à peu en concevoir et en adapter une qui puisse faire ressortir un ensemble d'articles pertinents pour notre analyse. Cela nous a notamment permis de soulever certains points non prévus comme la grande diversité des maladies incurables ayant déjà eu un lien avec l'intelligence artificielle. Ainsi, pour ne pas être débordé par le nombre d'informations nous nous sommes par la suite restreints aux 6 maladies incurables et aux deux types de maladies les plus présents dans nos recherches.

Il en a découlé l'équation de recherche suivante :

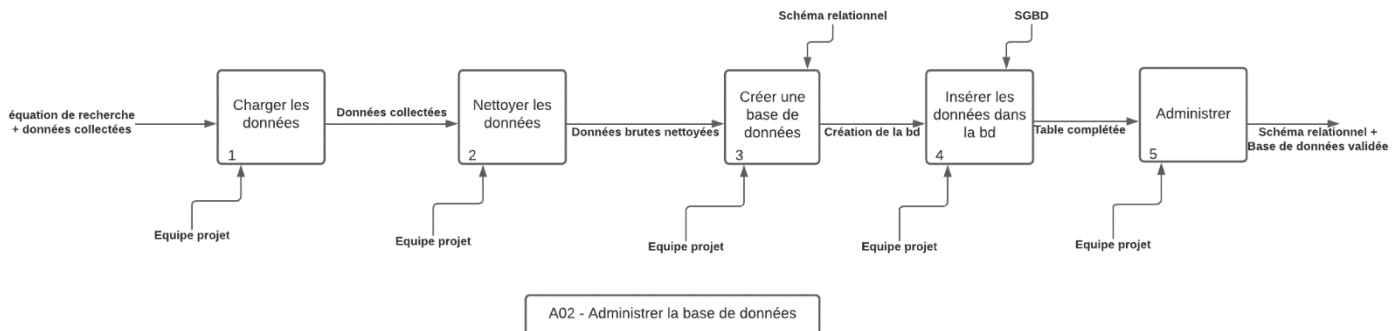
- (Artificial Intelligence or Machine Learning or Deep Learning or Neural Network or Big Data) and (Image Recognition Diagnostic or Detecting or Drug Design) and (Cardiovascular Disease or Cancer or Asthma or Alzheimer or Diabetes or AIDS or Schizophrenia or Neurodegenerative Diseases)

Une fois l'équation de recherche fixée et un nombre d'articles satisfaisant pour mener notre étude, nous nous sommes lancés dans l'extraction des articles depuis Python. Pubmed étant bien fournie, nous avons pu faire apparaître 11 000 résultats mais l'extraction n'a permis au final d'en récolter au maximum 10 000 offrant tout de même une gamme assez intéressante d'articles différents.

L'extraction ayant duré un temps non négligeable et au vu du temps estimé pour le nettoyage des données, nous avons décidé de stocker les articles récupérés au format brut dans un fichier csv et d'en extraire au fur et à mesure de nouvelles colonnes avec les informations qui nous seraient utiles (titre, nom des auteurs, date de publication ...).

C'est après avoir eu une idée plus claire des informations que l'on avait à disposition et la mise à jour finale du mcd que nous avons pu créer notre base de données et débiter le nettoyage des données.

## 2. Préparation des données



La partie pouvant avoir une forte influence sur les analyses graphiques est bien celle de la préparation des données. En effet, les données brutes récoltées au départ nécessitent un nettoyage tel que la suppression des balises html potentielles, la lemmatisation des mots ou encore la mise en minuscule des mots qui vont ainsi rendre les représentations graphiques plus pertinentes si le nettoyage est bien pensé par rapport aux informations que l'on souhaite analyser.

Ainsi, nous avons d'abord sélectionné les parties de l'article que nous voulions dans notre étude (le titre, les auteurs, le pays de l'étude, la date de publication, l'abstract, les mots-clés ainsi que la présence ou non dans le texte de maladies et/ou domaines en lien avec l'Intelligence Artificielle)

Un long travail de traitement a alors été fait pour ne pas fausser nos résultats, avec notamment :

La lemmatisation des textes en enlevant toutes les majuscules du texte, en supprimant les ponctuations, en étudiant les synonymes. À titre d'exemple, pour récupérer le pays il a fallu faire en sorte que « USA », « Usa », « U.S.A », « Etats Unis », « Etats-Unis », « New-York » ne fassent qu'une seule entité.

La restriction de certains termes, pour les maladies par exemple nous avons choisi les 6 maladies les plus courantes afin que nos graphiques soient pertinents en ne se basant pas sur les maladies très peu mentionnées (avec moins de 5 occurrences pour 10 000 articles)

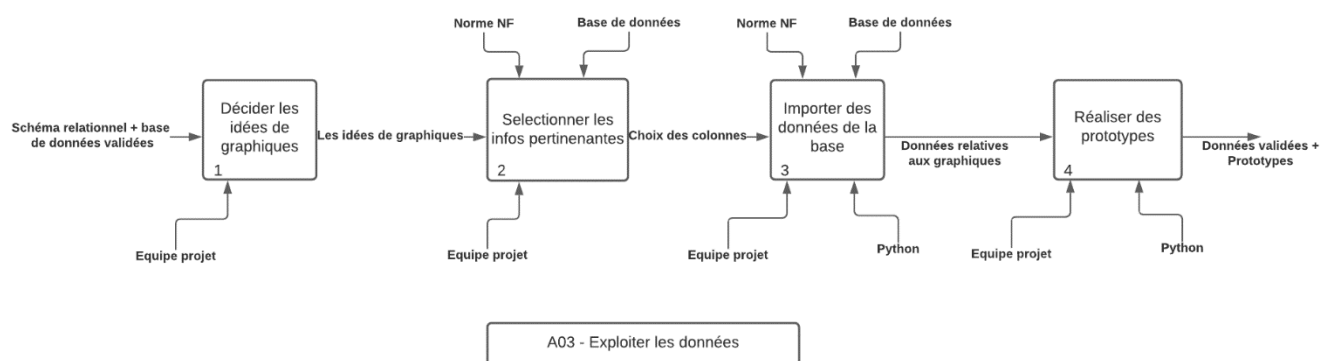
Les premières données nettoyées ont alors pu être insérées dans notre base ainsi que les termes choisis au départ manuellement (noms des maladies, des domaines).

Il s'en est suivi une deuxième phase plus poussée du traitement de texte. Pour poursuivre sur l'exemple précédent concernant les pays, nous nous sommes rendus compte que les études faites aux Etats Unis étaient les plus fréquentes, nous nous sommes alors dit qu'il

serait tout aussi intéressant de regarder la répartition de ces études dans les différents états des Etats Unis ce qui a nécessité un re-nettoyage de certaines colonnes.

Une fois cette deuxième phase de traitement finie nous avons pu implémenter notre base avec les données finales, prêtes à être analysées.

### 3. Valorisation des données



Le début de l'exploitation de données a commencé par la réflexion de graphiques que l'on souhaitait afficher dans notre dashboard.

Les idées qui ont été traitées sont :

- Le département, la ville et le pays, l'institut, l'université des auteurs de l'article
- L'occurrence des mots clés par domaine d'articles ou domaine de maladie
- Les domaines médicaux où l'IA est intervenue (la cancérologie, la dermatologie ...)
- La fréquence d'apparition des mots à partir d'un TF\_IDF
- Le nombre d'article au fil des années

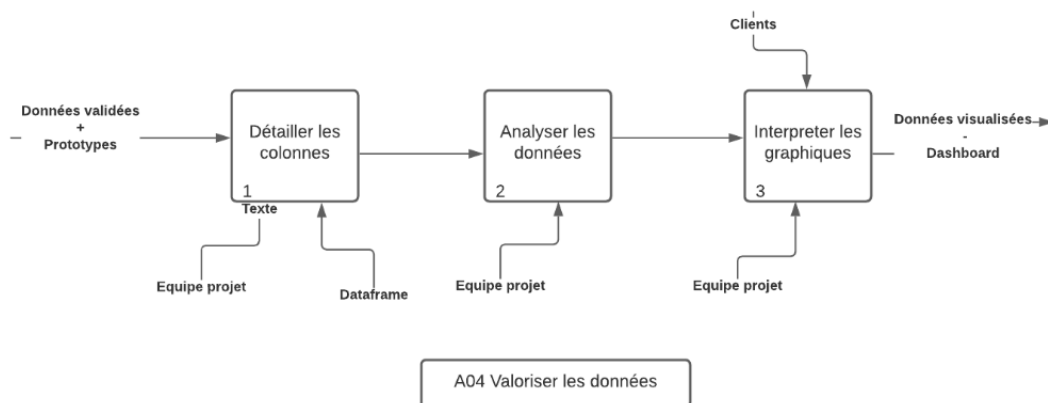
Toutes ces idées ont été codées et le code est à disposition dans les fichiers joints, cependant toutes n'ont pas été gardées pour la fin soit dû au fait que les graphiques ne faisaient apparaître aucune information pertinente ou alors que les données récoltées étaient très faibles (ce qui est arrivé pour l'université, le département ou le laboratoire des auteurs qui ont réalisés l'étude)

Après inspection des graphiques finaux que l'on souhaité conservé, on a dû garder comme colonne d'intérêt :

- La date de parution de l'article
- Les mots clés relevés
- L'identifiant des articles
- Le titre des articles
- Les pays à l'origine de l'article
- 

Le dataframe avec l'ensemble des colonnes même celles non conservées est également joint en fichier.

#### 4. Diffusion des connaissances



Comme expliqué précédemment, nous avons eu plusieurs idées mais certaines après une première analyse ne paraissaient pas pertinentes et n'ont donc pas été gardées. Au final les colonnes de notre dataframe qui ont été utilisées sont :

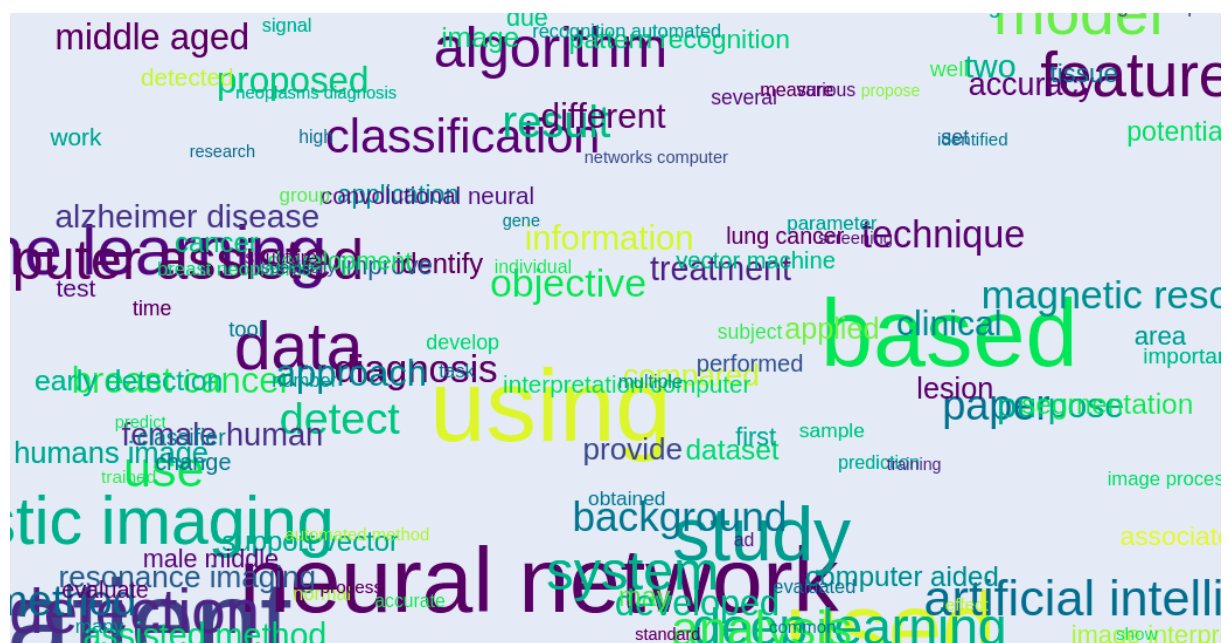
- Abstract pour compléter notre dictionnaire des mots clés
- Le titre de l'article
- Les mots-clés
- Le pays dans lequel les auteurs de l'article ont étudiés
- L'année de parution de l'article

Comme évoqué dans la partie précédente, nous avons utilisé l'algorithme de représentation textuelle TF\_IDF dans la partie abstract des articles qui nous a permis de relever de nouveaux mots clés pour nos analyses. On peut en citer quelques-uns :

- Pour les méthodes d'intelligence artificielle : imaging methods, deep learning, robotics, support vector machine, classification...
- Pour les domaines d'application : diagnostic, detection, drug design, patient care.

La première phase d'analyse a été dans l'objectif d'avoir une vue d'ensemble de nos articles, et de juger de leur pertinence.

## 1 – Le nuage des mots du corpus

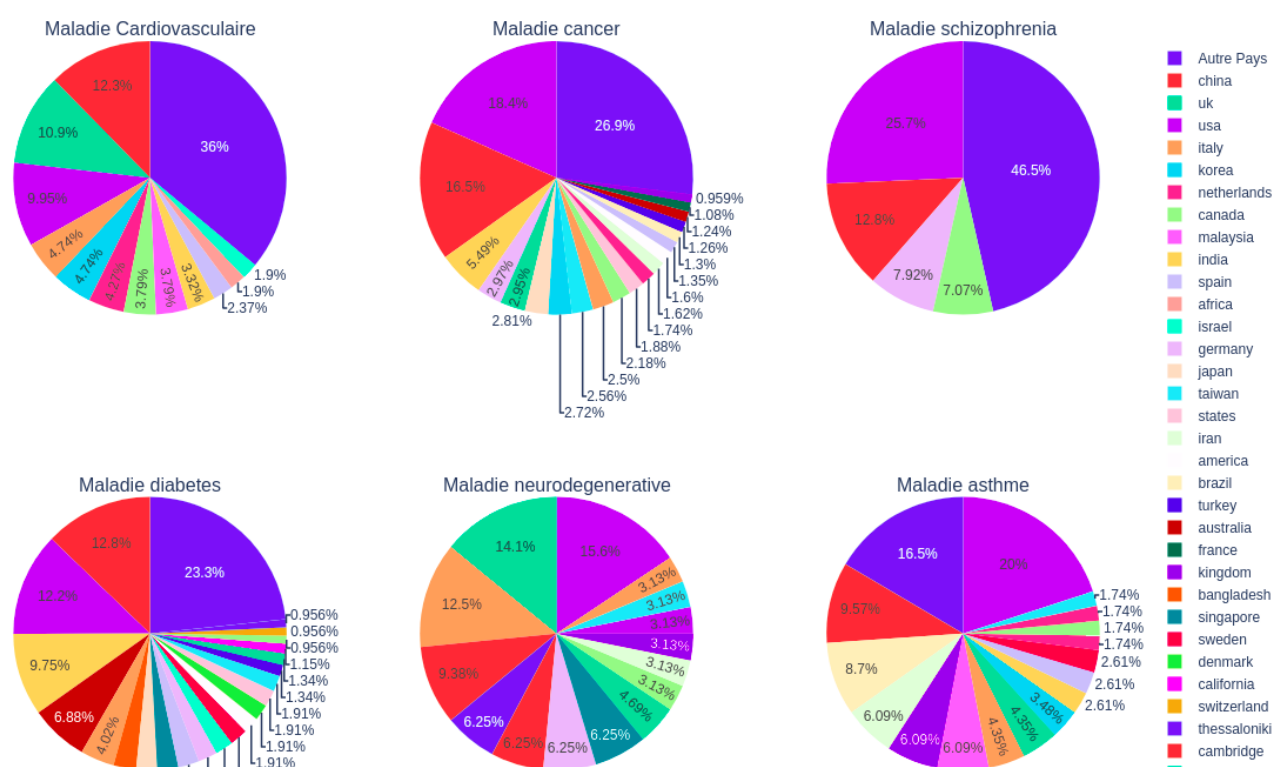


Ainsi, le premier nuage des mots présents dans l'ensemble de nos colonnes textuelles a mis en évidence trois types de vocabulaire :

- Les termes autour des méthodes d'intelligence artificielle tels que : Classification, neural network, algorthm, deep learning, data, prediction ...
- Ceux en lien avec les maladies : disease, alzheimer, clinical, cancer, gene ...
- Les domaines dans lesquelles l'IA intervient : early detection, diagnosis, treatment

Cela nous a donc assuré qu'une grande partie de nos articles liaient maladies incurables et IA et étaient donc bien pertinents pour notre sujet.

## 2 – Les pays de publication par maladie



Pour chacune des maladies, on peut remarquer plus ou moins une grande proportion de « autre pays » mais qui n'est pas vraiment pertinent ici car il s'agit seulement d'une accumulation d'un grand nombre de pays peu représentés.

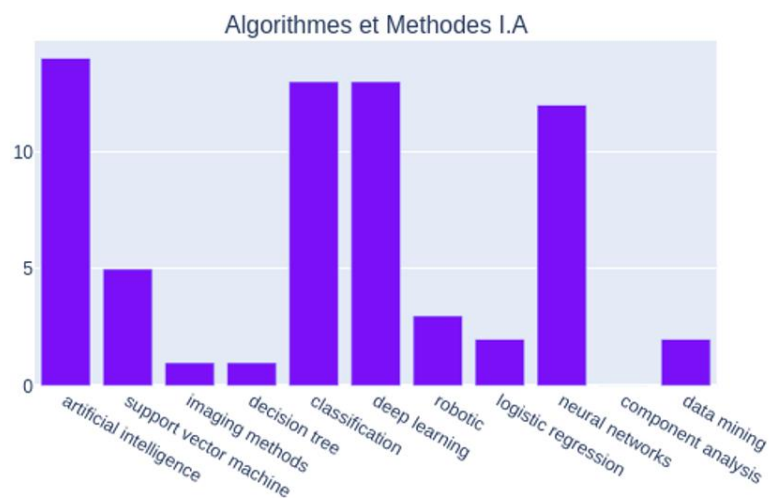
La tendance qui nous intéresse est celle qui arrive ensuite avec une majorité d'études chinoises et des États-Unis suivi de pays européens. Sans grande surprise il s'agit de pays où le niveau de recherche scientifique est très développé. Les articles paraissent donc assez pertinents pour être analysés.

On arrive maintenant aux graphiques plus complexes qui vont nous permettre d'apporter des réponses plus précises à notre sujet.

A noter que chacune des maladies était liée très différemment à l'intelligence artificielle. On a alors trouvé plus intéressant de faire des analyses par maladies et types de maladie plutôt que l'ensemble des maladies en général.

Pour chaque maladie on va retrouver 4 graphiques distincts. On va ici voir en détail le cas des maladies neurodégénératives. Les conclusions des autres maladies seront aussi apportées mais par soucis de lisibilité les graphiques correspondants seront en annexe.

### 3 – Les méthodes d'Intelligence artificielle utilisées



L'intelligence artificielle est un domaine très vaste, on peut par exemple y compter le machine learning, le deep learning et la robotique qui eux-mêmes présentent un nombre important de méthodes.

On a ainsi mis des mots clés de différents niveaux et ainsi vu si les méthodes mettaient en avant des méthodes très précises ou seulement des domaines.

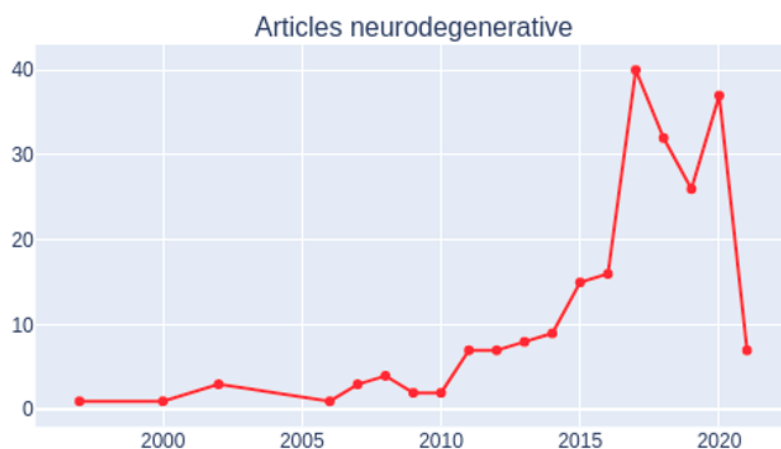
Exemple de 4 niveaux :

Artificial intelligence → Machine learning → Classification → Logistic Regression

Pour les maladies neurodégénératives, on remarque qu'il y a une forte présence d'algorithmes de classifications, de deep learning avec notamment les neural networks ce qui donne l'intuition que les maladies neurodégénératives observent une plus forte contribution des parties machine et deep learning de l'IA.



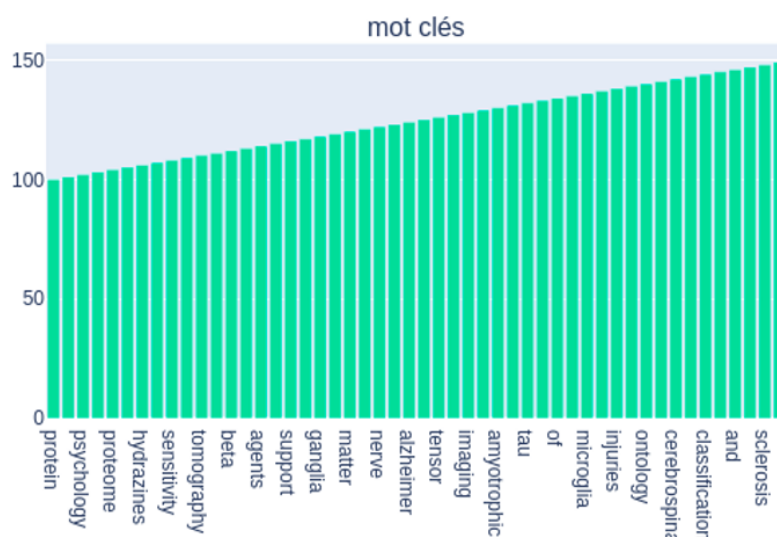
#### 4 – Evolution du nombre de parution d'articles



Sans grande surprise, on observe une augmentation chaque année du nombre d'articles liant l'IA et les maladies neurodégénératives, la dernière diminution s'explique par le fait que l'année 2021 n'est pas encore finie.

Cette augmentation d'articles au fil des années laisse supposer que la recherche scientifique dans ce domaine l'est aussi, notamment ces dernières années.

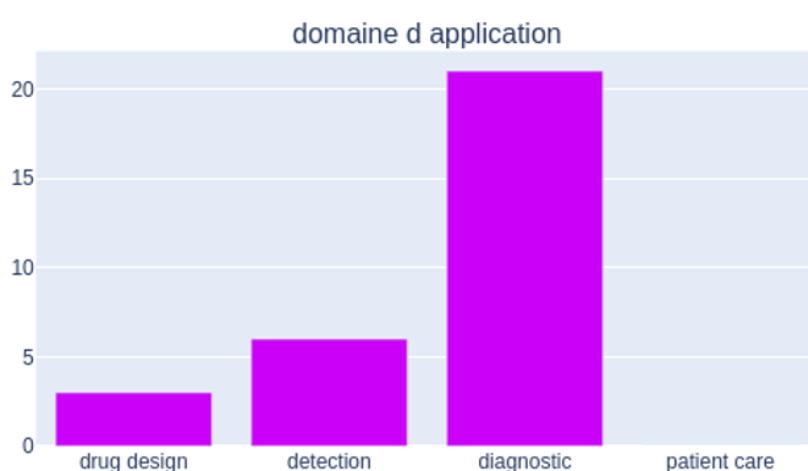
#### 5 – Fréquences des mots clés



On a tout à l'heure fait une première analyse des mots les plus récurrents dans le corpus de texte ce qui nous a permis d'abord de voir que les articles affichaient une corrélation entre l'étude de maladies incurables et de l'IA mais aussi d'enrichir notre vocabulaire de mots clés.

Nous avons procédé ici à la même analyse mais cette fois-ci en se basant seulement sur les sous-corpus spécifiques à chaque maladie. Dans le cas des maladies neurodégénératives on retrouve une majorité de mots distribués entre les lexiques de l'intelligence artificielle et le domaine médicale attestant de la contribution de l'intelligence artificielle.

## 6 – Evolution du nombre de parution d'articles



Durant nos recherches, on a trouvé que la contribution de l'intelligence artificielle dans le domaine médicale était partagée entre 4 secteurs :

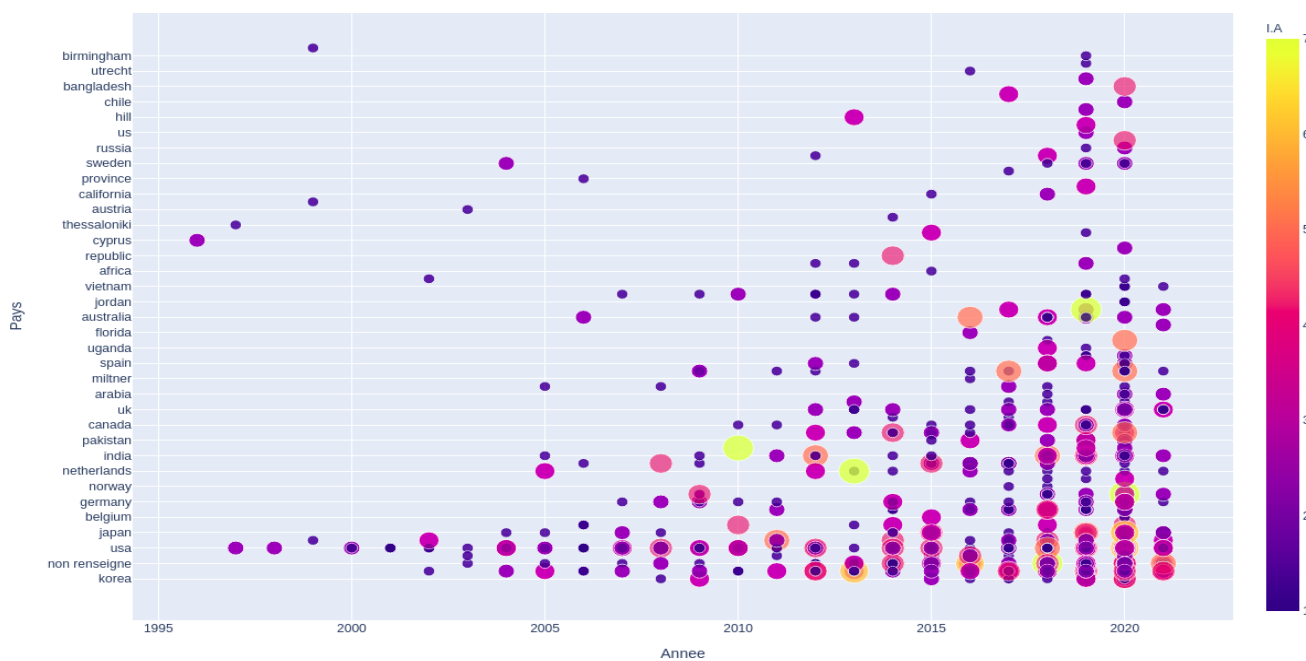
- La conception de médicaments, « drug design »
- L'aide à la décision médicale, « detection »
- Le diagnostic et dépistage d'un patient, « diagnostic »
- Le suivi du patient atteint d'une maladie, « patient care »

A partir de mots clés associés à chaque secteur, on a analysé la fréquence d'apparition de chacun d'entre eux. Pour les maladies neurodégénératives la contribution de l'intelligence artificielle semble plutôt se faire au niveau de la détection de la maladie, très peu de moyens semblent avoir été développés pour aider les personnes atteintes de ces maladies.

## 7 - Graphique regroupant l'ensemble des paramètres étudiés

Nous avons essayé de faire une analyse plus globale sur l'ensemble du corpus pour le dernier graphique, on prend en compte ici plusieurs paramètres :

- Le nombre d'articles par années en abscisse.
- Le pays de provenance des articles rédigées en ordonnée.
- La présence des méthodes d'intelligence artificielle : couleur qui varie du jaune (forte apparition) au violet (faible apparition).
- En pointant la souris sur un point on peut voir s'afficher les mots les plus présents dans l'article et ainsi détecter ceux liées à telle ou telle maladie.



Plusieurs informations nous sont données ici.

Premièrement on constate que pour chaque pays l'évolution a été la même : de plus en plus d'articles faisant référence à la contribution de l'intelligence artificielle dans les maladies incurables au fil du temps.

Cependant, on voit que cette avancée est différente en fonction des pays, en bas on retrouve des pays comme « USA », « Germany », « Japan », « India » qui globalement font ressortir les pays où l'on sait que la recherche scientifique est plus développée. Au contraire, au-dessus on va plutôt retrouver des pays comme le Bangladesh ou le Chili peu présent dans les recherches.

Également on s'aperçoit que les articles de plus en plus récurrents au fil des années sont aussi de plus en plus détaillés au sujet de la contribution de l'intelligence artificielle, les points les plus jaune-orangé se font voir en grande majorité dans les 10 dernières années.

# Gestion de configuration

## 1. Ressources matérielles

Chaque membre du groupe a pu travailler sur son ordinateur personnel, certains sous versions Windows et d'autres sous Linux mais cela n'a posé aucun problème de compatibilité. Également, pour tout la partie traitement des données, nous avons utilisés :

- ❖ Des logiciels pour traiter les données : Oracle SQL Développer et Jupyter Notebook.
- ❖ Un ensemble de bibliothèques Python : Pandas, Plotly, WorldCloud
- ❖ Un logiciel de planification : Gantt Project
- ❖ Les logiciels de communication : Zoom et Discord
- ❖ Les ordinateurs personnels : versions Linux et Windows

## 2. Les différentes versions exploitées

Comme expliqué dans la partie démarche et développement, nos idées de graphiques ont évolués au fur et à mesure du déroulement du projet, certaines idées ont été gardées d'autres non. En tout cas, tous ces changements nous ont menés à avoir plusieurs versions de code python. De même sorte pour la partie scrapping des données, nous étions partis sur un premier code qui n'a pas abouti car pas adapté à notre site web. Le code de scrapping retenu finalement étant disponible dans le fichier partagé.

Tout le long du projet on a eu à modifier notre base de données et ainsi eu plusieurs versions du modèle conceptuel des données (**MCD**). La version zéro ayant été créée avant de bien connaître le contenu des articles suivi de la version une où nous avons été guidés par M Mokadem qui a ensuite pu nous le valider.

Par la suite en fonction des nouvelles idées qui nous parvenaient nous avons ajouté une table concernant les différents **domaines d'application** de l'intelligence artificielle dans le traitement des maladies incurables. Nous avons également rajouté le **pays** dans lequel chaque **auteur** a effectué ses recherches. On a également essayé de préciser cette provenance pour certains pays en ajoutant un attribut **état** au départ pour préciser les différents états du pays majoritaire (les USA) et par la suite en précisant aussi les villes pour les études provenant

de France. Toujours dans cette même table on a modifié le **laboratoire** de l'**auteur** par le **département** auquel il est rattaché dû au peu de données concernant ces laboratoires. Enfin, on a supprimé l'attribut **brevet** de la table **article** car finalement non renseigné à l'intérieur et on en est ainsi arrivé au MCD final.

Un aperçu des MCD de première et dernière version :

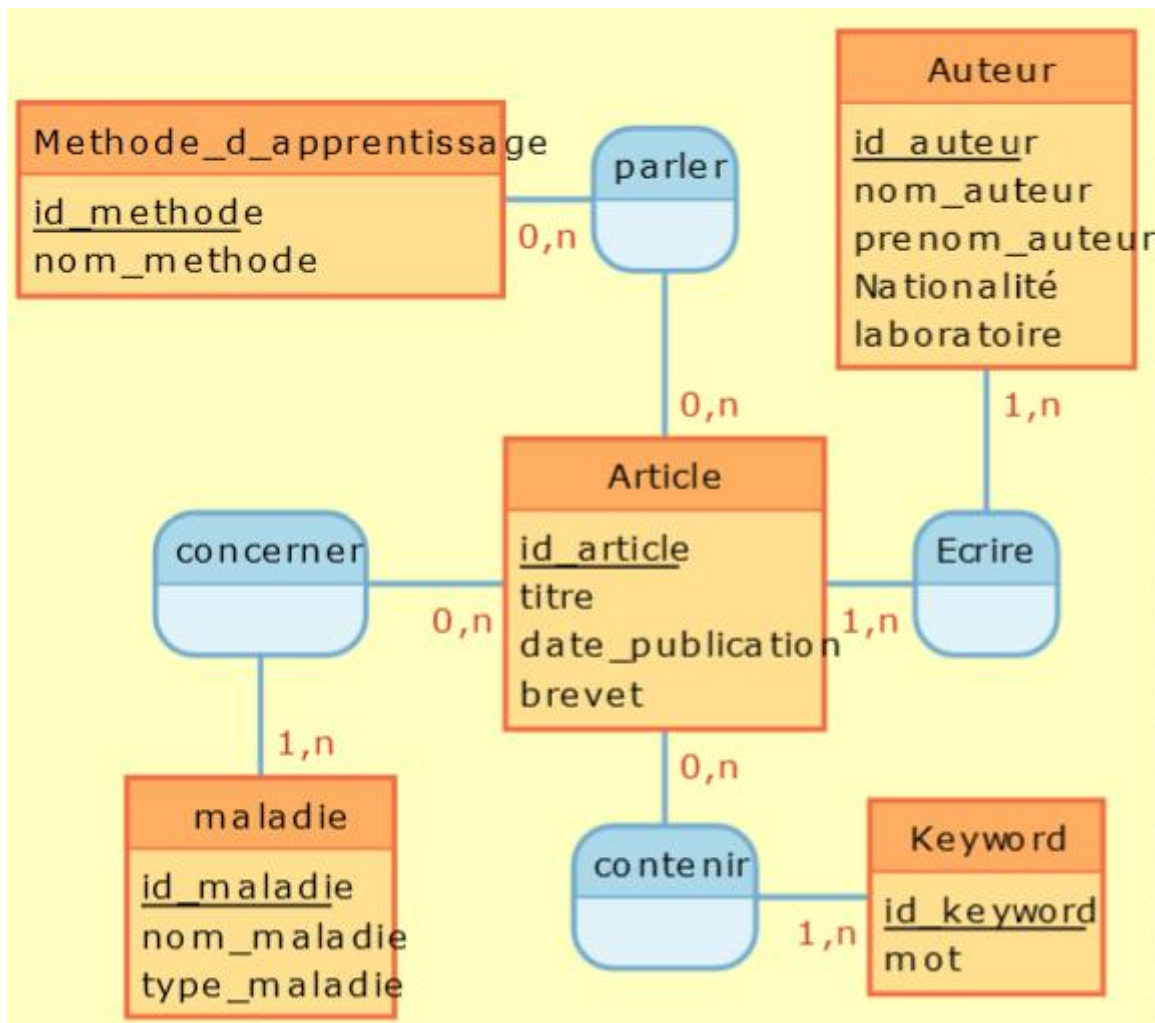


Schéma 1a – Modèle Entité-Association version 0

**METHODE\_D\_APPRENTISSAGE** ( id\_methode, nom\_methode )  
**PARLER** ( id\_methode, id\_article )  
**AUTEUR** ( id\_auteur, nom\_auteur, prenom\_auteur, Nationalité, laboratoire )  
**CONCERNER** ( id\_maladie, id\_article )  
**ARTICLE** ( id\_article, titre, date\_publication, brevet )  
**ECRIRE** ( id\_article, id\_auteur )  
**MALADIE** ( id\_maladie, nom\_maladie, type\_maladie )  
**CONTENIR** ( id\_article, id\_keyword )  
**KEYWORD** ( id\_keyword, mot )

Schéma 1b – Modèle Relationnel version finale

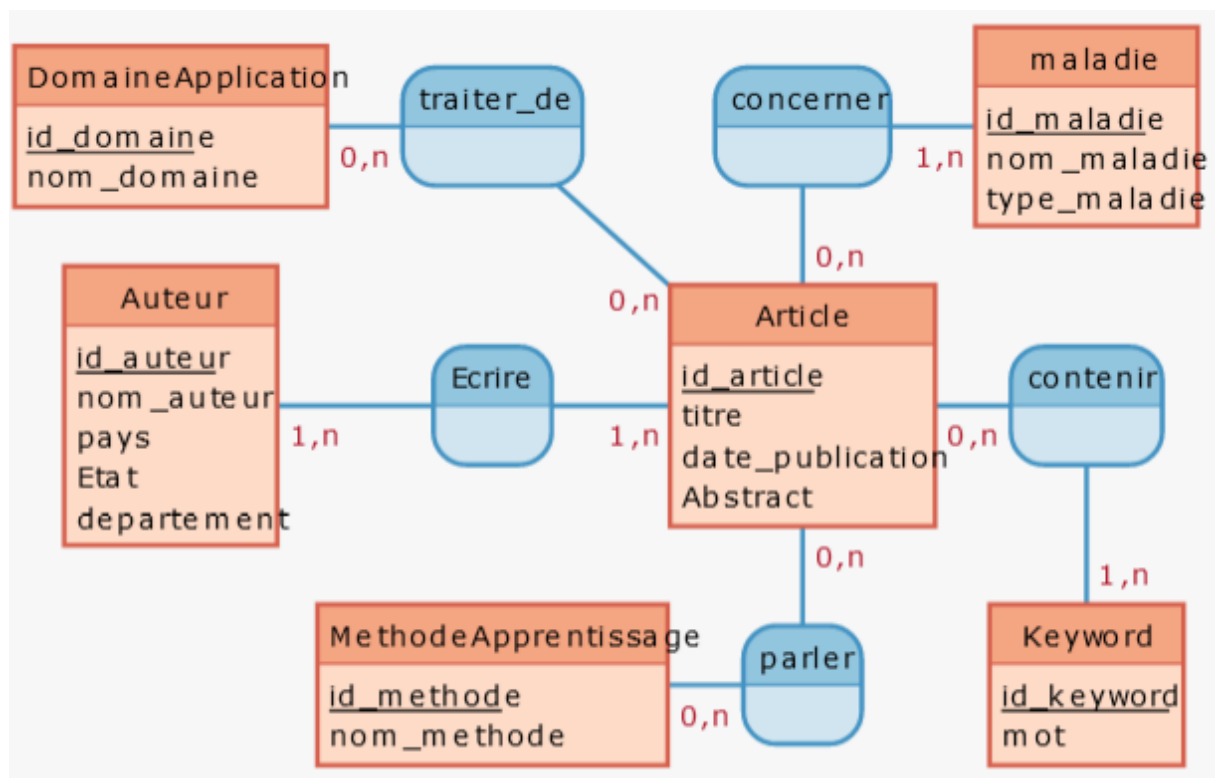


Schéma 2a – Modèle Entité-Association version finale

**DOMAINEAPPLICATION** ( id\_domaine, nom\_domaine )  
**TRAITER\_DE** ( id\_domaine, id\_article )  
**CONCERNER** ( id\_maladie, id\_article )  
**MALADIE** ( id\_maladie, nom\_maladie, type\_maladie )  
**AUTEUR** ( id\_auteur, nom\_auteur, pays, Etat, departement )  
**ECRIRE** ( id\_article, id\_auteur )  
**ARTICLE** ( id\_article, titre, date\_publication, Abstract )  
**CONTENIR** ( id\_article, id\_keyword )  
**METHODEAPPRENTISSAGE** ( id\_methode, nom\_methode )  
**PARLER** ( id\_methode, id\_article )  
**KEYWORD** ( id\_keyword, mot )

Schéma 2b – Modèle Relationnel version finale

# Assurance qualité

## 1. Revues

- **25/01/2021**

- **Objectif :**

Choisir un sujet et chercher les articles pertinents

- **Problèmes rencontrés :**

Un sujet trop peu précis au départ avec un nombre d'articles trouvés dépassant les 10 millions.  
Notre source de données n'était pas définie, Google Scholar et PubMed étaient tous les deux intéressants à exploiter.

- **Pour la semaine suivante :**

Réfléchir à une problématique plus précise et décider sur la source à utiliser.

- **01/02/2021**

- **Objectif :**

Valider notre sujet.

Concevoir une première équation de recherche et commencer à l'enrichir.

Se faire une idée sur les requêtes à effectuer.

- **Problèmes rencontrés :**

Les requêtes proposées sont peu affinées, nous n'atteignons pas le nombre d'articles voulu.

- **Pour la semaine suivante :**

Optimiser l'équation de recherches avec des synonymes.

- **08/02/2021**

- **Objectif :**

Avoir une première version V0 du MCD.

- **Pour la semaine suivante :**

Mettre à jour le MCD.

- **17/02/2021**

- **Objectif :**

Valider le MCD.

Créer la base de données V0 correspondante



- **22/02/2021**
  - **Objectif :**
    - Commencer le scrapping
  - **Problèmes rencontrés :**
    - Le code et les fonctions utilisées durant le projet inter-promo ne fonctionnent pas sur notre site de scrapping
  - **Pour la semaine suivante :**
    - Avancer dans le scrapping
    - Mettre à jour le MCD et la base de données et avoir la version finale
- **27/02/2021**
  - **Objectif :**
    - Avancer dans le scrapping
    - Commencer le dashboard
  - **Pour la semaine suivante :**
    - Finir le scrapping
    - Préparer les idées des graphes
    - Peupler la base de données
    - Faire les requêtes SQL
- **29/02/2021**
  - **Objectif :**
    - Commencer les graphiques et faire les analyses
  - **Pour la semaine suivante :**
    - Avancer dans les graphiques
    - Finaliser le dashboard
- **05/03/2021**
  - **Objectif :**
    - Analyser les graphiques
  - **Pour la semaine suivante :**
    - Avancer dans les graphiques
    - Finaliser le dashboard
    - Mettre à jour le rapport et le diagramme de gantt
- **12/03/2021**

- **Objectif :** Mettre en propre
- **Pour la semaine suivante :** Finaliser le rapport

# Conclusion et bilan du projet

## 1. Conclusion

La problématique de base était d'étudier la contribution de l'intelligence artificielle au sein des maladies incurables, pour éviter de s'éparpiller dans un grand nombre de maladies nous sommes restreints aux plus courantes. Cela a tout de même permis de déceler un lien qui s'est raffermi au fil des années avec des articles de plus en plus nombreux liant ces deux sujets.

On a réussi à mettre en avant les 4 principales thématiques de l'intelligence artificielle dans le domaine médical :

- La conception de médicaments
- L'aide à la décision médicale
- Le diagnostic et dépistage d'un patient
- Le suivi du patient atteint d'une maladie

En étudiant les maladies de plus près on remarque tout de même des tendances de certaines méthodes ou domaine en fonction des maladies, le suivi médical plus présent pour le diabète ou la détection de maladies plus présentes pour les maladies neurodégénérative.

## 2. Bilan

Dans sa globalité nous sommes assez fiers de la tournure du projet, il s'agissait d'un travail conséquent qui était endurci par le distanciel mais nous avons réussi à aller au bout de notre objectif qui était de fournir un Dashboard conforme au cahier des charges.

Cependant, en décortiquant un peu le déroulement on a pu faire face à pas mal de difficultés. La partie la plus décisive du projet a été la récolte des articles, chose à laquelle nous n'avions pas été formés en cours et qui était différente durant le projet interpromo ce qui nous a mené à nous plonger dans une façon de coder nouvelle et de s'y renseigner.

Un point évident est le fait que tout s'est déroulé en distanciel, le fait de travail ensemble mais chacun chez soi emmène une toute autre ambiance de travail qui peut vite devenir agréable même si pour le coup nous nous étions bien organisés ayant déjà travaillés de nombreuses fois ensemble sur zoom et discord, nous avons réussi à garder un bon rythme de travail

Lors du peuplement de la base nous avons croisé un souci lié à l'espace de stockage, le problème a été résolu assez vite étant donné que l'on a directement relevé l'information mais cela nous a quand même bloqué une journée sur la dernière semaine qui est très importante où on ne pouvait pas tester entièrement nos nouveaux graphiques sur la base de données.

Un point qui n'est du ressort de personne a été qu'un membre du groupe a eu des complications personnelles qui ont menées à réattribuer certaines tâches au cours du projet.

Néanmoins le projet est resté très formateur, c'est en appliquant réellement ce que l'on apprend que l'on développe nos compétences. Ici on a pu approfondir les différentes connaissances acquises durant notre formation (la programmation informatique, l'administration d'une base de données, les analyses statistiques, la gestion de projet et le traitement de l'information) et aussi une occasion pour pratiquer et transmettre entre nous les connaissances que nous avons acquis durant les projets inter promo

Nous avons pu aussi développer nos compétences en traitement de l'information, en appliquant des graphiques et donc des perspectives d'analyses différentes.

S'ajoute à ces compétences professionnelles les compétences personnelles qui sont forcément mise en jeu et développées dans un projet en équipe basé sur la force de proposition, l'entraide et avant tout, la communication.

On a ainsi chacun pu partager ce que l'on a appris durant les projets interpromo et élargir nos connaissances dans le domaine technique, projets où chacun se voit étudier un aspect différent de notre formation.

Les projets interpromo ont apporté une très grande aide, l'avantage de notre groupe a été que chaque phase du projet avait déjà été traitée par un membre lors d'un des deux projets interpromo. Également dès le départ nous avons adopté une méthode de travail similaire où avant chaque phase nous allions nous documenter sur internet pour consolider les outils que l'on aurait à utiliser.

Nous avons aussi déjà pour habitude de travailler ensemble ce qui a facilité la communication et l'échange. Ainsi, malgré quelques différences de points de vue sur quelques tâches durant le projet, nous avons réussi à accorder nos envies pour aboutir à une solution de qualité qui satisfait l'ensemble des membres ce qui a contribué à harmoniser l'efficacité de l'équipe.

Ce projet nous a permis de nous immerger dans un univers professionnel et nous a donc apporté à chacun une idée sur l'organisation dans le monde professionnel ce qui nous permettra de nous adapter plus facilement lors de notre stage.

# ANNEXE

Le code du dashboard est disponible dans le fichier partagé, voici tout de même le rendu final :



## Projet Dashbord M1 SID

la contribution de I.A dans les maladies incurables

Premieres Graphes et Analyse

Cardiovasculaire & IA

**Cancer & IA**

Schizophrenia & IA

Diabetes & IA

Deurodegenerative & IA

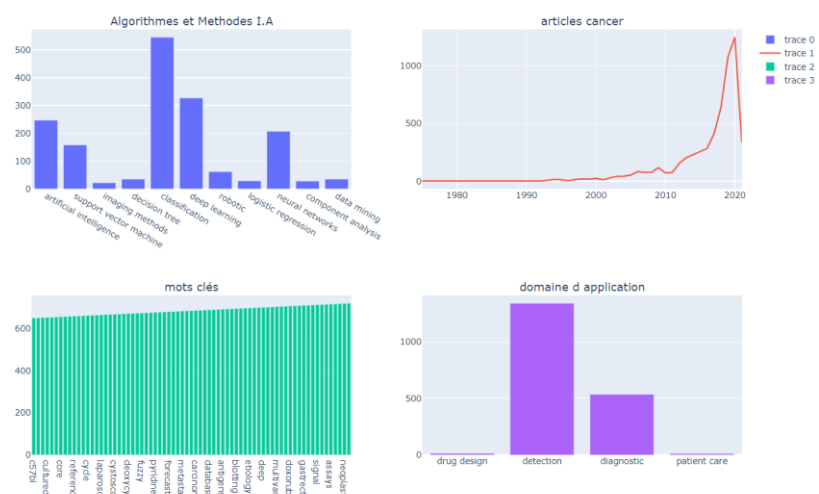
Asthma & IA

Alzheimer & IA

Algorithmes IA et Maladies

Nuage des mots totale

## Cancer



## Projet Dashbord M1 SID

la contribution de I.A dans les maladies incurables

Premieres Graphes et Analyse

Cardiovasculaire & IA

Cancer & IA

**Schizophrenia & IA**

Diabetes & IA

Deurodegenerative & IA

Asthma & IA

Alzheimer & IA

Algorithmes IA et Maladies

Nuage des mots totale

## Schizophrenia



## Projet Dashbord M1 SID

la contribution de I.A dans les maladies incurables

Premieres Graphes et Analyse

Cardiovasculaire & IA

Cancer & IA

Schizophrenia & IA

**Diabetes & IA**

Deurodegenerative & IA

Asthma & IA

Alzheimer & IA

Algorithmes IA et Maladies

Nuage des mots totale

## Diabetes



## Projet Dashboard M1 SID

la contribution de I.A dans les maladies incurables

Premieres Graphes et Analyse

Cardiovasculaire & IA

Cancer & IA

Schizophrenia & IA

Diabetes & IA

**Deurodegenerative & IA**

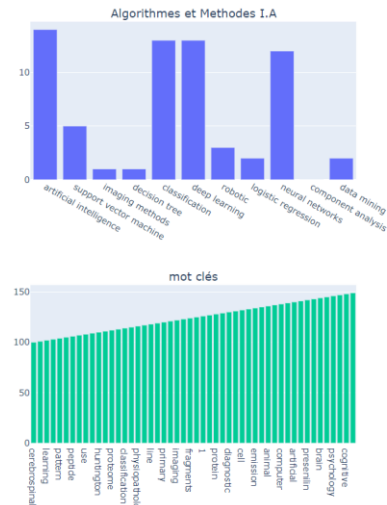
Asthma & IA

Alzheimer & IA

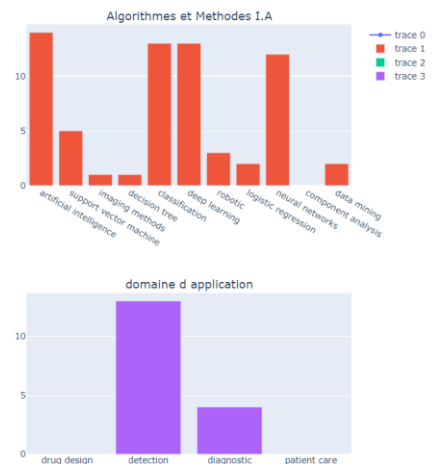
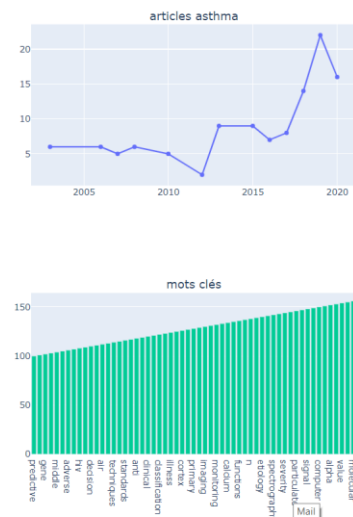
Algorithmes IA et Maladies

Nuage des mots totale

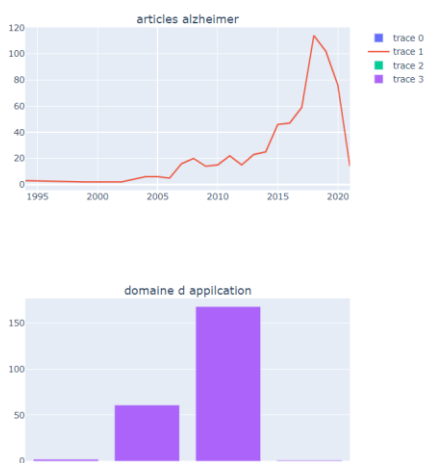
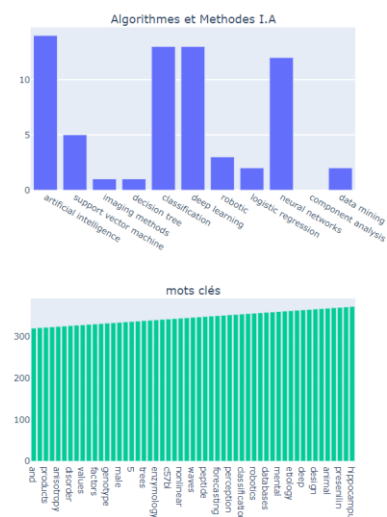
## Neurodegenerative



## Asthma



## Alzheimer



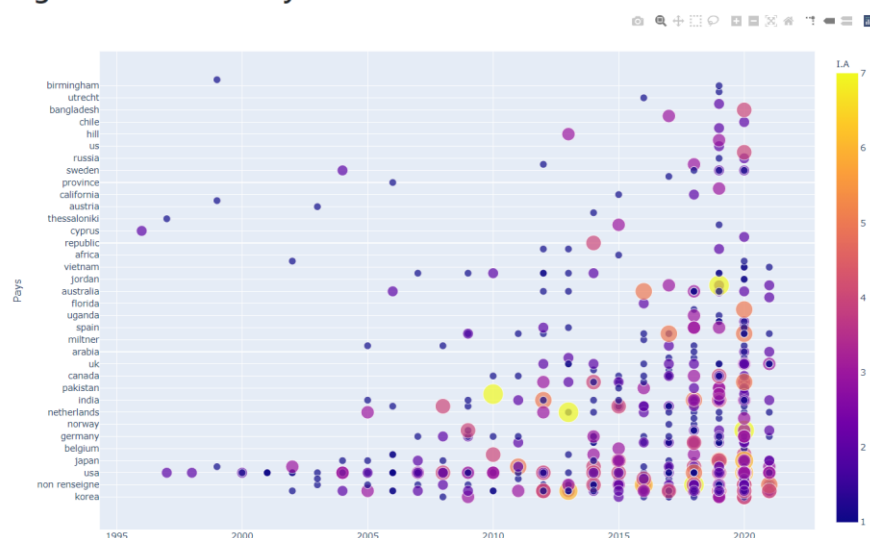


Projet  
Dashbord  
M1 SID

la contribution de IA dans les maladies incurables

- Premieres Graphes et Analyse
- Cardiovasculaire & IA
- Cancer & IA
- Schizophrenia & IA
- Diabetes & IA
- Deurodegenerative & IA
- Asthma & IA
- Alzheimer & IA
- Algorithmes IA et Maladies**
- Nuage des mots totale

Graphe générale des articles des maladie en fonctions des années, des algorithmes I.A et Pays



Projet  
Dashbord  
M1 SID

la contribution de IA dans les maladies incurables

- Premieres Graphes et Analyse
- Cardiovasculaire & IA
- Cancer & IA
- Schizophrenia & IA
- Diabetes & IA
- Deurodegenerative & IA
- Asthma & IA
- Alzheimer & IA
- Algorithmes IA et Maladies**
- Nuage des mots totale

Nuage des mots dans notre scrapping abstracts

