



Mathématiques : moteur de l'intelligence artificielle

Le Big Data – des données à exploiter en classe.

La demande initiale a pour objectif de travailler le sujet de l'intelligence artificielle (IA). De nos jours, et qui plus est devant le phénomène induit par des applications telles chatGPT ou Midjourney, il devient impossible de faire l'impasse sur cette notion, qu'il s'agisse du traitement massif de données accessibles en ligne ou sur les suggestions des services de streaming par exemple.

Les académies de Strasbourg, Aix-Marseille et Rennes ont travaillé sur ce thème au cours de l'année 2021-2022 autour de la création de jeux à objectifs mathématiques et de stratégie de résolution de jeu.

L'académie de Nantes a fait le choix d'approcher l'IA via le big data, c'est-à-dire l'utilisation d'un nombre important de données. L'idée était aussi de proposer un renouvellement des approches possibles autour des statistiques, notions présentes dans les programmes de la sixième à la terminale. En effet, il est nécessaire d'obtenir des données exploitables et, si possible, en quantité suffisamment importante pour établir des prévisions. L'organisation adoptée se définit en 3 axes de travail :

- ✓ création de données par les élèves ;
- ✓ proposition de pistes, traitement et exploitation de données ;
- ✓ utilisation des techniques apprises pour résoudre des problèmes riches et intéressants.

1) Créer des données avec les élèves.

Lorsqu'un professeur souhaite traiter le chapitre autour des statistiques, il doit utiliser des données. Une approche possible est de les créer puis de les exploiter. Pour les obtenir, il est possible de partir du quotidien des élèves. Cependant, un travail en amont autour de la pertinence des questions est indispensable. Il permet en effet d'anticiper le traitement à réaliser une fois les données établies et de faire comprendre aux élèves certains biais éventuels. Par exemple une personne peut-elle répondre plusieurs fois à un questionnaire au risque d'influencer trop fortement les résultats ?

Pour créer ces données, plusieurs méthodes sont possibles allant de l'interrogation de personnes ([sondage classique en classe](#)), à l'utilisation de réseaux sociaux (sondage sur twitter [saison1](#) – [saison 2](#)), à l'exploitation de données récoltées pendant une activité (descente de skieur – géolocalisation en SNT).

2) Créer des activités à partir de données existantes – s’interroger sur les données.

Les données récoltées à partir des élèves peuvent s’avérer insuffisantes, inexploitable ou moins riches que prévues. A l’image de ce qui se fait dans d’autres matières (SES – SVT – etc), les professeurs de mathématiques doivent pouvoir utiliser des données recueillies sur différents sites. Plusieurs questions apparaissent :

- ✓ où trouver des données brutes c’est à dire non traitées ?

Certains sites apportent de telles données. L’utilisation d’un tableur est très souvent nécessaire.

Exemples : [site de l’insee](#), [données publiques françaises](#).

- ✓ Comment identifier la signification des informations contenues dans les fichiers ?

Le professeur est amené à aider à la compréhension de ces données. Ainsi, l’étude de la population d’un département permet de mettre en place certains indicateurs statistiques de position au collège ([population en Sarthe](#)).

Le professeur peut aussi exploiter les données pour n’en retenir qu’une partie et cibler le travail ([répartition des richesses / maths complémentaires](#)).

- ✓ Comment avoir un regard critique sur les données traitées ?

Le professeur peut inciter ses élèves à tester la limite des valeurs obtenues. Les élèves peuvent alors proposer une extrapolation pour compléter les nombres initiaux. Le choix du modèle est alors intéressant pour confirmer ou infirmer une hypothèse ([population en Mayenne](#)).

3) Faire traiter des données brutes par les élèves – utiliser les fonctionnalités du tableur.

En mathématiques, le professeur peut faire découvrir aux élèves comment traiter des données brutes, obtenues à partir d’un fichier ([csv par exemple](#)). En effet, un fichier de données brutes est un fichier difficile à lire sans un minimum de traitement. Comment rendre alors lisible une liste de caractéristiques, parfois longue de plusieurs centaines de lignes ? Dans de nombreuses activités, nous utilisons très souvent en premier lieu le tri du tableur (objets perdus). Mais d’autres fonctionnalités, plus inhabituelles sont aussi pertinentes à faire découvrir : par exemple les fonctions GAUCHE et DROITE, NB.SI ou CONCATENER ([vérifier la loi de Benford](#)).

Il est à noter que, avec l’arrivée de la programmation dans les programmes, les élèves ont souvent moins manipulé le tableur qu’auparavant. [Des séances d’introduction peuvent donc être utilisées](#). Elles permettront par la suite d’exploiter des fichiers plus importants ([élections présidentielles aux Etats-Unis](#) - [étude démographique des départements](#) - [étude démographique de villes](#)).

4) Automatiser l'exploitation de données avec un langage de programmation.

Parfois certains fichiers s'avèrent très conséquents avec des milliers de lignes. Ils sont donc extrêmement compliqués à analyser pour des collégiens ou des lycéens. Si un tableur fournit de très nombreux outils, ils ne sont pas toujours compréhensibles pour la classe. La formation est aussi insuffisante pour pouvoir exploiter efficacement ces outils : nommer une plage de données importantes n'est pas un réflexe attendu ni chez les enseignants, ni chez les élèves. Il est alors pertinent d'habituer les élèves à l'utilisation d'outils complémentaires à un tableur. Ainsi, l'exploitation du langage scratch est une possibilité offerte au collège ([stage en entreprise - températures à Bâle](#)), tout comme la programmation en python l'est pour le lycée (La catastrophe du Titanic).

5) De la recherche de données à la prise de décision – le rapprochement avec l'intelligence artificielle.

L'intelligence artificielle utilise des données afin de répondre à une question. En classe, les élèves sont parfois mis en situation de suivre un cheminement similaire les amenant à mobiliser les connaissances acquises pour répondre à une problématique. Les deux exemples ci-dessous illustrent cette possibilité :

[Exemple 1](#), à partir de l'extrait suivant, trouvé sur le site de Loire Atlantique : « *le niveau moyen de la mer à Saint-Nazaire a augmenté de 20 cm sur les 200 dernières années, les scénarios les plus pessimistes des experts du Groupe Intergouvernemental d'Experts sur l'évolution du Climat (GIEC) prédisent une élévation de 60 cm du niveau marin d'ici à 2100* ». Cette affirmation a été soumise à des élèves de 4^{ème}. Pour y répondre, les élèves ont travaillé sur différents fichiers. Tout d'abord, ils ont analysé les données de l'année 2020. Puis, ils ont exploité celles récupérées depuis ...1960 (60 années disponibles) ! Pour terminer, ils ont pu utiliser un autre outil pour visualiser les moyennes (GeoGebra) et ils ont dû choisir un modèle pour conclure (droite d'ajustement). Ce travail motivant a permis de mieux comprendre le travail d'un chercheur.

[Exemple 2](#), à partir d'un défi vieux de plus d'un siècle : « *l'histoire commence à Nenana en Alaska, en 1917, un groupe d'ingénieurs explorateurs pour le chemin de fer parient 800 US\$ en devinant quand la rivière se brisera sous le poids d'un trépied. Depuis, ce petit jeu est devenu une tradition qui se poursuit encore aujourd'hui* ».



Cette question est d'abord posée aux élèves. Chacun propose alors spontanément une date. Puis, en SNT et sans se référer à la question initiale, le professeur met en place un parcours qui débute par un rappel autour du tableur avant de proposer des exercices de plus en plus élaborés autour de l'étude des données. Ce travail permet de retourner aux données proposées sur le site du défi et d'affiner les réponses ...ou pas.

Conclusion :

L'étude et l'exploitation du big data est une première étape dans la compréhension d'un mode de fonctionnement de l'IA.

Plusieurs points peuvent se dégager de nos différentes expériences :

- éviter des données hors sols ;
- utiliser davantage les sites data ;
- mettre du sens aux indicateurs statistiques ;
- *Prendre davantage conscience de l'apport du tableur par rapport à la calculatrice ;*
- stimuler l'apprentissage des statistiques en se donnant des objectifs ambitieux.

Nous partageons les ressources produites autour des Big Data en vue de contribuer à une meilleure appréhension de l'IA en tant que telle.