

Thème 03 - Une histoire du vivant

Chapitre 05 - L'intelligence artificielle

1 Traitement des données informatiques

● Définition d'un ordinateur

On appelle **ordinateur** un système capable de traiter de l'information selon un programme écrit dans un langage informatique, et comportant au minimum un **processeur** et une **mémoire vive** (Fig. 1).

● Les données informatiques

Les données informatiques que peut traiter un ordinateur sont de natures diverses : textes, sons, images mais également programmes informatiques. Elles sont toujours numériques, ce qui nécessite parfois un processus de **numérisation** (Fig. 2).

Exemple : La norme internationale ASCII définit la façon de coder les caractères d'un texte. Chaque caractère y est codé par un octet (valeur binaire codée sur 8 bits). La valeur 1000001 sera attribuée à A, 1000010 à B, etc. La taille du fichier (en nombre d'octets) est alors égale au nombre de caractères du texte.

Les programmes peuvent être transformés en fichiers exécutables qui ont l'extension .exe sous le système d'exploitation Windows ou .app sous Mac OS.

Les données informatiques sont stockées sur divers supports (disques durs, clés USB, etc.) et peuvent être transportées sur un réseau informatique (dont internet).

2 Les programmes informatiques

Pour élaborer des programmes, les programmeurs utilisent des langages informatiques traduits ensuite par la machine en suites de 0 et de 1.

Un programme informatique peut comporter jusqu'à plusieurs centaines de millions de lignes de code. La probabilité d'erreurs de codage (appelées **bogues** ou **bugs**) est importante. Cela peut parfois avoir de graves conséquences.

Il existe de nombreux types de processeurs. Certains langages de programmation sont spécifiques à un type de processeur et sont très difficiles à utiliser.

Les **langages de haut niveau** sont indépendants du matériel utilisé et beaucoup plus simples à utiliser et à déchiffrer, donc à déboguer*.

Exemples : Les langages Java, Python ou Scratch sont des langages de haut niveau.

Autour de nous, on trouve de nombreux ordinateurs programmés par des programmeurs dans les sociétés qui les commercialisent et/ou des ordinateurs programmables directement par l'utilisateur grâce à des interfaces simples à utiliser (Fig. 3).

3 L'intelligence artificielle

● Définition

Le terme d'**intelligence artificielle** (IA) désigne un ensemble de théories et de techniques de programmation dont le but est de simuler l'intelligence humaine, que ce soit dans la capacité d'apprentissage, les perceptions ou la prise d'initiatives.

Exemples : Les résultats de recherches en intelligence artificielle permettent aujourd'hui à des ordinateurs de reconnaître des formes et des visages, de conduire une voiture, de traduire un texte, etc. (Fig. 4).

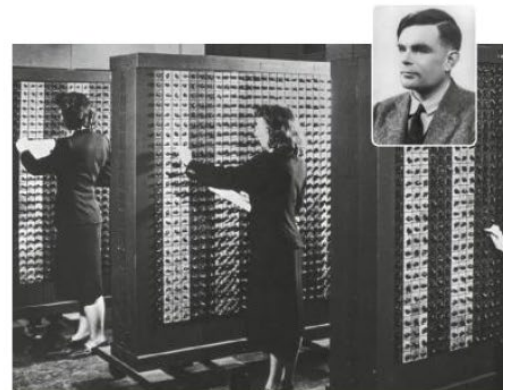


Fig. 1 : C'est le mathématicien Alan Turing (1912-1954) qui imagine le concept d'ordinateur moderne. L'ENIAC, qui pèse plus de 30 tonnes, est le premier ordinateur construit selon ses principes (1945).



Fig. 2 : Grâce à un scanner, des textes ou images peuvent être convertis en données informatiques, c'est-à-dire en données numériques.



a Puce de GPS.

Fig. 3 : L'utilisateur peut **(b)** ou non **(a)** interagir avec l'ordinateur.



b Thermostat réglable.

Les recherches en IA n'en sont cependant actuellement qu'à leur début, aucun ordinateur actuel ne pouvant être qualifié encore de véritablement « intelligent ».

● L'apprentissage machine

L'apprentissage machine permet aux ordinateurs de réagir dans des situations inédites à partir de données d'entraînement sélectionnées par le programmeur.

Des méthodes mathématiques de repérage de tendance (corrélations, similarités) et de probabilités sont utilisées sur de grandes quantités de données (*big data*). Elles permettent ensuite d'effectuer des prédictions ou de prendre des décisions dans des situations nouvelles.

La qualité et la représentativité des données d'entraînement sont essentielles pour la fiabilité des réponses.

Si les données d'entraînement comportent un **biais**, les résultats donnés par l'ordinateur seront de mauvaise qualité, ce qui peut avoir de graves conséquences.

Certains usages de l'IA peuvent ainsi poser des problèmes éthiques. Des biais dans l'apprentissage pourraient par exemple entraîner des discriminations.

Exemples : Des algorithmes utilisant l'apprentissage machine peuvent être utilisés pour sélectionner automatiquement des CV pour un recruteur, calculer une probabilité de radicalisation pour la police ou de maladie pour une assurance.

● L'inférence bayésienne

L'inférence bayésienne est une méthode de calcul de probabilités de causes à partir des probabilités de leurs effets. Elle sert notamment à modéliser des relations au sein de systèmes complexes.

Exemples : L'inférence bayésienne peut servir à se prononcer sur la crédibilité de diagnostics dans les domaines médicaux et industriels ou l'analyse de spams.

Dans un test médical, si la fréquence de la maladie dans la population générale est f_m , alors le nombre de personnes réellement malades s'écrit (Fig. 5) :

$$\frac{f_m \cdot \text{fréquence(VP)}}{f_m \cdot \text{fréquence(VP)} + (1 - f_m) \cdot \text{fréquence(FP)}}$$



Fig. 4 : Spot, le chien robot de la société Boston Dynamics. Doté d'un bras dorsal, il est déjà utilisé pour explorer de façon autonome des zones dangereuses pour l'Homme.

Test	Malade	Sain
Positif	vrai positif (VP)	faux positif (FP)
Négatif	faux négatif (FN)	vrai négatif (VN)

Fig. 5 : Exemple de tableau de contingence utilisé dans l'approche bayésienne.

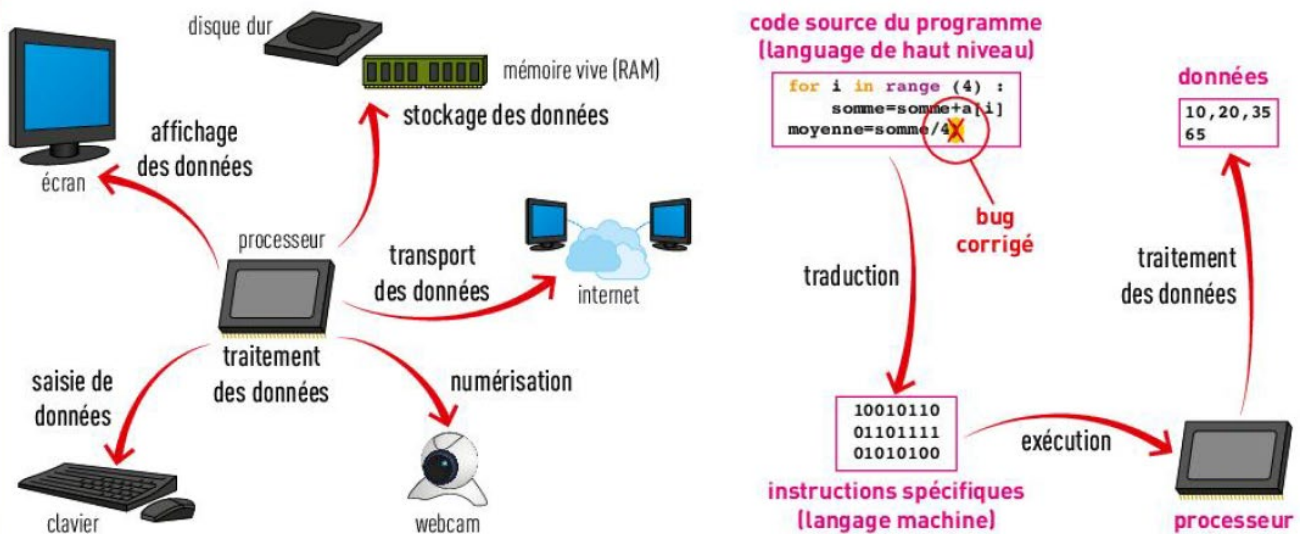
Le vocabulaire à retenir

- **Apprentissage machine :** technique qui permet à un ordinateur d'apprendre à réagir dans des situations nouvelles à partir de données d'entraînement entrées par un programmeur.
- **Biais :** dans l'apprentissage machine, erreur systématique dans les prédictions due à des données d'entraînement mal choisies.
- **Bogue (ou bug) :** erreur de codage entraînant un comportement de l'ordinateur aberrant.
- **Inférence bayésienne :** à partir de faits observés, cette méthode permet de définir la probabilité de causes possibles.
- **Intelligence artificielle :** techniques de programmation informatique visant à imiter l'intelligence humaine.
- **Langage de haut niveau :** langage informatique conçu pour être facilement compréhensible et utilisé par l'être humain.
- **Mémoire vive :** mémoire informatique présente dans un ordinateur et permettant de stocker temporairement des données.
- **Numérisation :** conversion d'une information (texte, image, etc.) en données numériques.
- **Ordinateur :** machine de traitement d'information comprenant au moins un processeur et une mémoire vive.
- **Processeur :** partie d'un ordinateur qui traite les données et exécute les instructions des programmes.
- **Tableau de contingence :** représentation de données dans un tableau selon deux critères, l'un en ligne, l'autre en colonne.

Résumé

1

L'ordinateur : un système capable de traiter des données

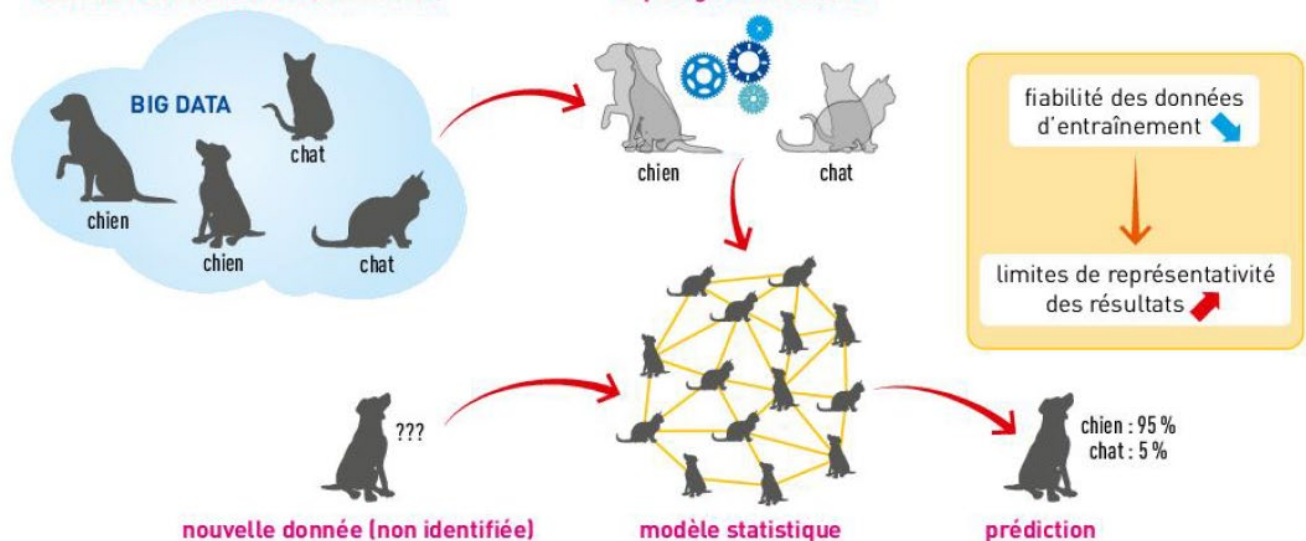


2

L'apprentissage machine

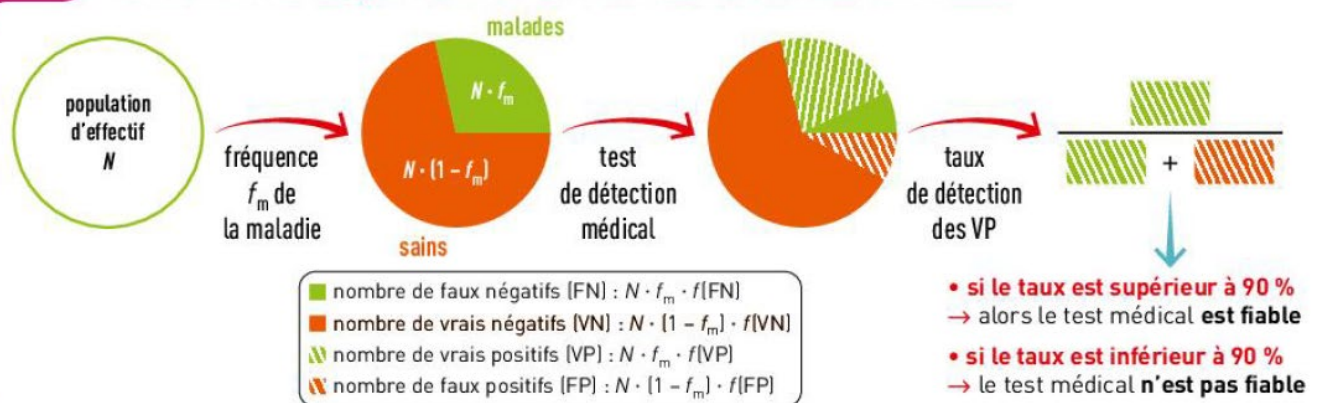
données d'entraînement (identifiées)

repérage de tendances



3

L'inférence bayésienne : test de détection d'anomalie



Exercices



Pour chaque question, indiquer la (ou les) bonne(s) réponse(s).

	a	b	c	d
1 Au niveau de son processeur, l'ordinateur traite l'information sous forme :	d'un programme de haut niveau.	d'un fichier texte.	de séries de 0 et de 1 (bits).	de langage naturel.
2 En programmation, un bug :	est un dysfonctionnement des composants de l'ordinateur.	est une erreur de logique ou de syntaxe dans un algorithme.	peut aboutir à un résultat incohérent.	engendre systématiquement un message d'erreur.
3 L'intelligence artificielle :	utilise des données d'entraînement pour en extraire un modèle applicable à de nouvelles données.	progresses de manière autonome sans intervention humaine.	est une méthode d'enseignement de l'informatique par des robots.	est basée sur des algorithmes élaborés pour effectuer une tâche complexe.

4 Vrai ou Faux ?

Indiquer si les affirmations suivantes sont vraies ou fausses.

- « Un fichier son d'un morceau de musique est un fichier exécutable. »
- « Une IA bien entraînée est fiable à 100 %. »
- « Augmenter le nombre de données d'entraînement augmente la fiabilité d'une IA. »
- « Une IA ne peut remplacer l'humain que sur des tâches précises où elle est entraînée ; elle ne peut pas apprendre d'elle-même sans entraînement complémentaire. »

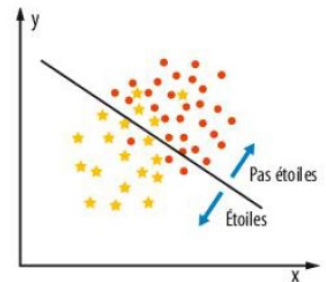
5 Phrases à construire

Construire une phrase à l'aide des mots suivants.

- pixel - octets - couleur - image
- instruction - informations - exécutable - type
- algorithme - coder - IA - modèle - programmation - entraîner

7 Faux positifs

L'objectif est de sélectionner les étoiles bleues mais pas les ronds rouges. On fixe une droite de séparation. Tous les points situés sous la droite sont considérés par le système comme étant une étoile.



1. Remplir le tableau de contingence suivant :

	Étoile	Rond
Bien classifié		
Mal classifié		

2. Calculer le taux de faux positifs dans cette situation.

8 Repérer des erreurs dans un programme

Le programme en Python ci-contre est censé calculer la longueur d'onde correspondant au pic d'émission lumineuse d'un corps noir en fonction de sa température (loi de Wien). Bien que les relations mathématiques soient correctes, ce programme ne fonctionne pas car il comporte des bugs (ou bogues).

- Repérer et corriger les bugs empêchant l'exécution de ce programme.
 - Le fichier contenant ce code source a pour nom « lambdamax.py ».
- Déterminer s'il s'agit d'un fichier exécutable.

```
#programme permettant de calculer
#la valeur de λmax (en nm) du spectre
#émis par un corps noir

def convertit_kelvin(tempCelsius)
    tempKelvin==tempCelsius+273,15
    return tempKelvin

def calcule_lambda_max(tempCelsius)):
    tempKelvin=convertit_kelvin(tempCelsius)
    cste=2897700
    lambdaMax==cste/tempKelvin
    return lambdaMax
```


9 Du modèle à son exploitation

Dans une master-class de jazz, on demande aux participants combien de temps ils ont répété un morceau avant de passer devant le jury.

Heures de répétition	Note	Heures de répétition	Note
20	5	5	9
40	20	8	8
30	15	26	14
15	10	32	16
25	11	35	17
22	12	0	2
2,5	6		



1. **Observer** graphiquement s'il y a un lien entre le temps de travail et la note.
2. **Déterminer** l'équation de la courbe de tendance, ici une droite (on la donnera sous la forme $y = ax + b$), avec y représentant la note obtenue d'un participant en fonction du nombre d'heures x de répétition.
3. Lors de la sélection ultérieure, les participants donnent leur temps de répétition. **Prédire** si le musicien réussira la sélection ou pas (on considérera qu'il a échoué pour une note en dessous de 10).
4. Suite à l'examen, on observe si les participants ont été sélectionnés ou non. **Comparer** la prédiction au résultat réel pour remplir un tableau de contingence.
5. **Discuter** la pertinence du modèle pour la prédiction de la réussite.

Participant	Heures de répétition	Prédiction
Marcus	10	échec
Pauline	20	succès
Liam	10	échec
Ginko	30	succès

10 Différents symptômes pour une même erreur

Algorithme

Lorsqu'un développeur rédige un programme dans un langage donné, un de ses objectifs est de créer le programme le plus adapté à la finalité souhaitée. Par exemple, le gestionnaire d'une clinique vétérinaire souhaite compter plus facilement le nombre d'animaux présents dans son centre de soin. Il a l'idée de créer une liste L dans laquelle chaque terme correspondra au nombre d'animaux dans une des cages. Il crée ensuite une fonction `total_animaux(L)` qui calcule la somme des termes de cette liste :

```
def total_animaux(L):
    S=0
    for k in range(5):
        S=S+L[k]
    return S
```

Listes

```
L1 = [8, 4, 2, 0, 3]
L2 = [5, 2, -3, 6, 0]
L3 = [1, -2, 6, 3, 2, 4]
L4 = [4, 2, 9, 5]
```

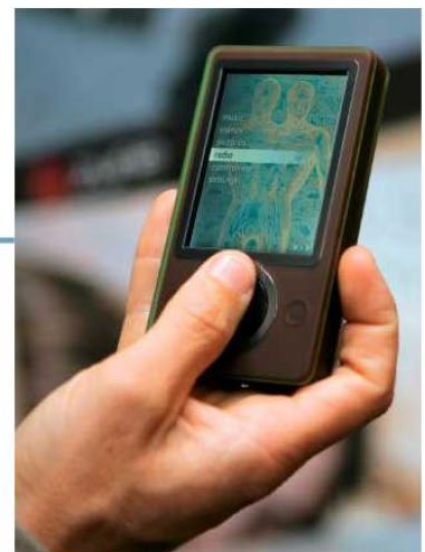
1. **Tester** cet algorithme avec les listes proposées. **Observer** les différentes réponses et **comparer** aux résultats calculés.
2. **Proposer** une explication pour expliquer ces différences.
3. **Proposer** une modification de cet algorithme afin qu'il fonctionne dans tous les cas.

13 Le bug du Zune

Le Zune était un appareil de Microsoft censé concurrencer l'iPod de Apple. Le dernier jour de 2008, année bissextile, tous les appareils commercialisés ont soudainement cessé de fonctionner. Le lendemain, après avoir rechargé la batterie, les appareils ont de nouveau pu fonctionner. Le code incriminé avait pour objectif de déterminer l'année en cours en comptant le nombre de jours écoulés depuis le début de l'année. L'algorithme était bloqué dans ce que l'on appelle une boucle infinie. Le code était exécuté perpétuellement mais sans jamais rien produire.

1. **Décrire** ligne par ligne l'extrait de l'algorithme de Zune.
2. **Tester** cet algorithme « à la main » avec le 242^e jour d'une année non bissextile, avec le 345^e jour d'une année bissextile.
3. **Expliquer** pourquoi il ne peut pas fonctionner le 366^e jour d'une année bissextile et proposer une modification.

```
Année= année en cours
Jour= numéro du jour en cours
Tant que (jour > 365) :
    Si Année est bissextile :
        Si Jour > 366 :
            On enlève 366 à Jour
            On augmente Année de 1
        Sinon :
            On enlève 365 à Jour
            On augmente Année de 1
```



14 Test de primalité

Prépa E3C

L'étude des nombres premiers joue un rôle très important en arithmétique et en cryptographie. On souhaite ici réaliser un programme qui permet de tester si un nombre est premier ou non.

Les programmes suivants comportent chacun une erreur.

1. Créer un jeu de données test et **anticiper** « à la main » les résultats.

2. **Exécuter** les programmes à l'aide du jeu de données et **identifier** puis **corriger** les bogues éventuels.

```
def primalite_Lucas(n):
    for k in range(2,n):
        if n%k==0:
            return False
        else:
            return True
```

Programme A

```
def primalite_Jules(n):
    for k in range(1,n):
        if n%k==0:
            return False
        return True
```

Programme B

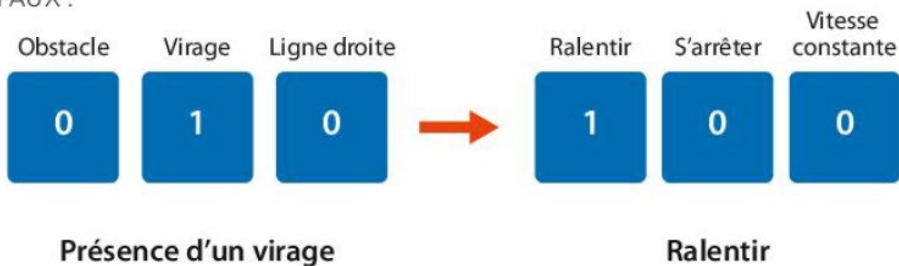
```
def primalite_Evariste(n):
    for k in range(2,n+1):
        if n%k==0:
            return False
        return True
```

Programme C

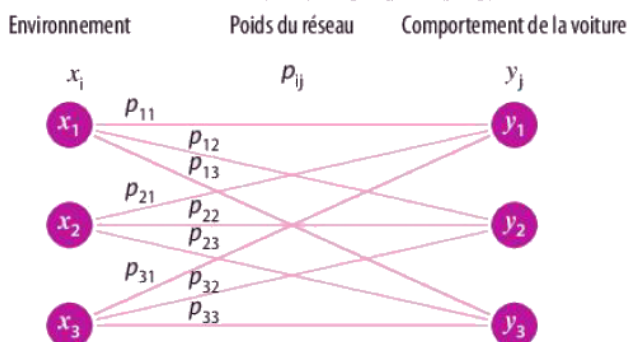
15 Entraîner une voiture autonome

On souhaite construire un algorithme capable d'adapter le comportement d'une voiture (ralentir, s'arrêter, maintenir la vitesse), en fonction de différentes situations (obstacle, virage libre, ligne droite libre). On considérera que les situations sont exclusives.

On représente cela par des triplets de bits, avec '1' signifiant VRAI et '0' signifiant FAUX :



- À partir de cet exemple, **modéliser** le comportement de la voiture pour les deux autres situations possibles.
- On imagine à présent un réseau liant 3 valeurs en entrée à 3 valeurs en sortie tel que :
On a par exemple $y_1 = x_1 \cdot p_{11} + x_2 \cdot p_{21} + x_3 \cdot p_{31}$.



À l'aide du comportement modélisé précédemment, **calculer** les valeurs manquantes dans le tableau.

Poids	Valeur
P_11	0
P_12	
P_13	
P_21	1
P_22	
P_23	
P_31	0
P_32	
P_33	

- Indiquer** si pour une situation non prévue : un panneau stop dans un virage libre, on peut utiliser le modèle précédent.
- Représenter** de même un réseau permettant de choisir entre deux activités (pique-nique ou cinéma) en fonction de la météo (soleil, pluie).

Exercice type

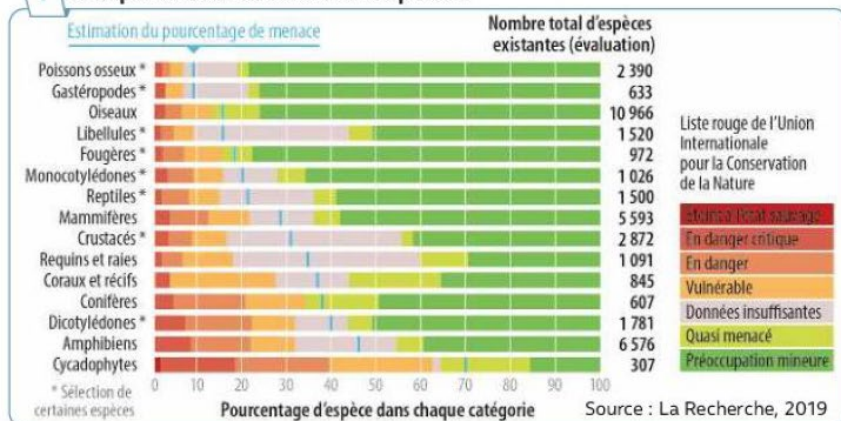
Épreuve commune de contrôle continu



Biodiversité, évolution et actions humaines (10 points)

La planète Terre que nous connaissons est l'héritage d'une longue histoire géologique et de l'évolution de la vie. L'Union Internationale pour la Conservation de la Nature (UICN), une organisation non gouvernementale, a réalisé un inventaire mondial de la biodiversité. On constate que l'abondance des espèces dans les populations terrestres a diminué de 25 % depuis la Préhistoire. L'être humain agit parfois directement sur la biodiversité. Connaître la biodiversité, actuelle et passée, est nécessaire pour agir et garantir l'avenir du vivant.

1 Risque d'extinction des espèces



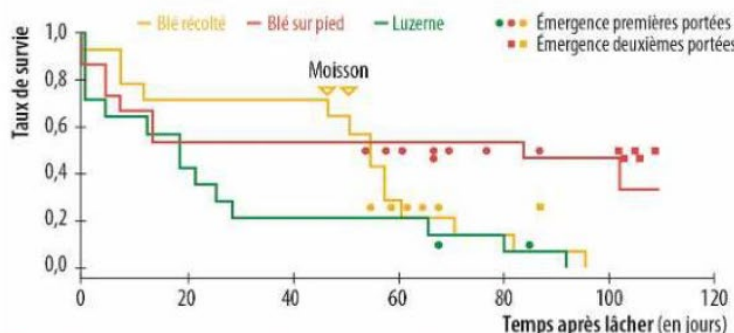
1. a. Si l'on regroupe les catégories « éteint à l'état sauvage », « en danger critique » et « en danger », le nombre d'espèces en danger est-il plus grand chez les Crustacés ou chez les Coraux des récifs ? Justifier la réponse.

b. « Moins de 10 % des espèces de poissons osseux sont menacées. Le risque d'extinction des espèces n'est donc pas si grave dans ce groupe animal ». Utiliser vos connaissances scientifiques pour commenter cette phrase.

2 La protection du Hamster d'Alsace

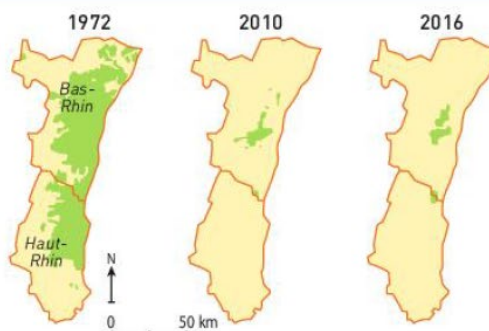


Le Grand Hamster (*Cricetus cricetus*) est un rongeur sauvage d'Alsace qui a obtenu le statut d'espèce protégée en 1993. Son habitat typique correspond aux zones de cultures fourragères et céréalières qui lui apportent gîte et ressources alimentaires. Les populations de cette espèce ont fortement diminué au cours du xx^e siècle, du fait des pratiques agricoles modernes, mais aussi de la lutte contre ces rongeurs (pesticides et piégeage) et du développement de l'urbanisme et des infrastructures routières. Afin d'estimer son effectif, 208 individus adultes ont été capturés, marqués puis relâchés en juin 2015. Deux mois plus tard, une nouvelle capture de 205 individus adultes a révélé que 43 d'entre eux étaient marqués.



b Taux de survie des femelles Grand Hamster lâchées dans différents environnements de culture

Source : ONCES, 2011



a Aire de répartition du Grand Hamster en Alsace entre 1972 et 2016

Source : ONCFS, 2018

Cause de mortalité (en %)	Total : 321 hamsters
Renard	31
Mustélide	24
Buse / rapace	18
Chat / chien / rat	8
Maladie	5
Travaux agricoles	3
Cause indéterminée	11

c Principales causes de mortalité de hamsters

Source : Ruiters, 2011

2a. Nommer la technique de comptage des individus utilisée en 2015. Utiliser les données de l'analyse pour calculer l'effectif du Grand Hamster d'Alsace.

2b. Comparer le taux de survie et le nombre de générations de Grand Hamster selon le type de culture. Expliquer.

3 Des sanctuaires pour préserver la biodiversité

À l'échelle mondiale actuelle, 14,8 % des terres et 7,3 % des océans sont des aires protégées. D'après des chercheurs du laboratoire d'écologie alpine de Grenoble, en protéger seulement 5 % de plus permettrait de tripler la variété des espèces protégées. Ainsi, 3 743 espèces supplémentaires d'oiseaux et de mammifères seraient sauvegardées.

Pour définir des sites optimaux de protection de la biodiversité, ces chercheurs utilisent un modèle s'appuyant classiquement sur la richesse en espèces. Ils utilisent aussi deux autres indicateurs : la diversité évolutive et la diversité fonctionnelle.

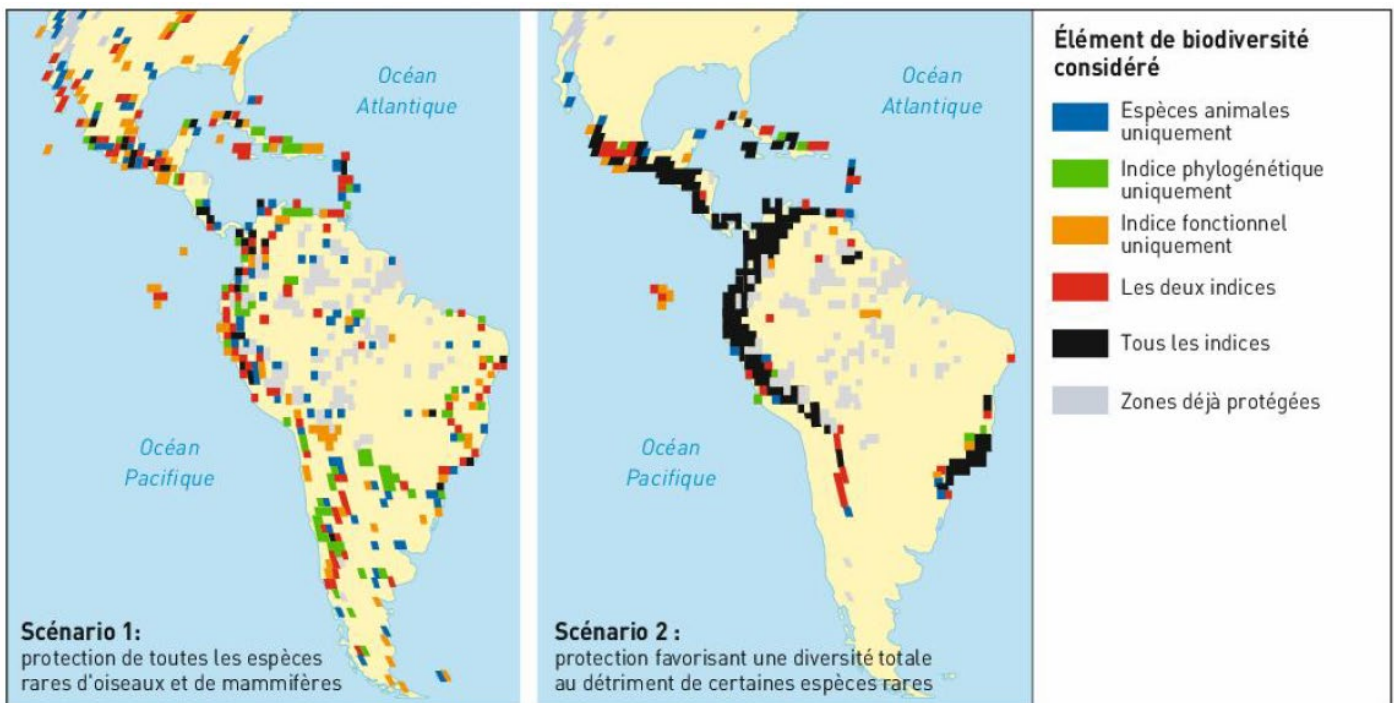
La diversité évolutive prend en compte l'importance de l'histoire évolutive d'une espèce. Par exemple, la protection de l'Ibis géant, considéré comme « en danger critique

d'extinction » par l'UICN, préserverait à elle seule 57 millions d'années d'histoire évolutive, soit la somme de toutes les années d'évolution qui ont été nécessaires pour arriver à cette espèce.

La diversité fonctionnelle, elle, représente l'ensemble des fonctions écologiques remplies par les espèces. Ainsi, un écosystème composé d'herbivores, de carnivores, d'insectivores, d'animaux nocturnes et diurnes, de pollinisateurs présente une grande diversité fonctionnelle.

La question des stratégies à adopter se pose : agir au niveau mondial ou au niveau local ? Protéger toutes les espèces ou donner la priorité aux espèces endémiques (localisées dans un secteur géographique restreint) ? Les chercheurs ont modélisé deux scénarios.

Résultat de deux simulations restituant les zones à protéger pour favoriser la protection de la biodiversité, dans l'hypothèse où on augmenterait au niveau mondial les aires protégées de 5 %



Source : B. Robert, La Recherche, 2020

Dans les faits, cependant, classer une aire ne garantit pas sa réelle protection. En effet, les zones les plus riches en biodiversité sont généralement situées dans les pays les plus pauvres où les gardes, trop peu nombreux, ne peuvent

assurer la surveillance de l'immensité des parcs nationaux concernés. De plus, des zones protégées subissent des pressions financières liées à leur exploitation.

3. Le document indique que « protéger seulement 5 % de plus permettrait de tripler la variété des espèces protégées. Ainsi, 3 743 espèces supplémentaires d'oiseaux et de mammifères seraient sauvegardées. »

a. Calculer le nombre d'espèces sauvegardées avant l'augmentation de 5 % des zones.

b. Si à chaque fois que l'on protège 5 % de surface en plus, on triple la variété des espèces protégées, calculer combien d'espèces seraient protégées après deux augmentations successives de 5 %.

4. À l'aide des connaissances et du doc 3, indiquer les forces évolutives en jeu dans la partie de phrase « 57 millions d'années d'histoire évolutive, soit la somme de toutes les années d'évolution qui ont été nécessaires pour arriver à cette espèce ».

5. En vous appuyant sur les documents 2 et 3, rédiger un texte pour résumer la réflexion actuelle pour agir en faveur de la biodiversité. Relever les difficultés éventuelles de ces actions.