

# POUR BIEN COMMENCER

## Quelques notions déjà vues

### ES 1<sup>re</sup> Coder le son

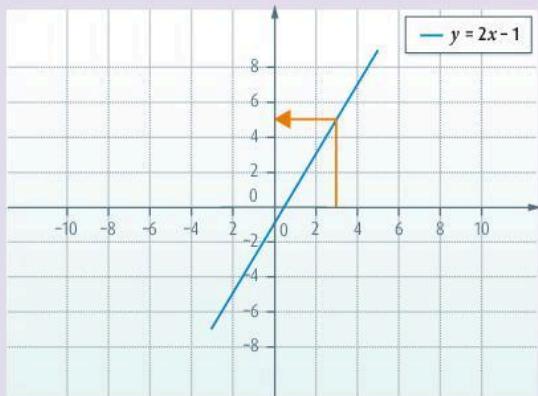
Avant d'être utilisable par un ordinateur, toute information doit être numérisée et convertie en binaire. Ceci est valable pour du texte, une image, de la vidéo ou du son, comme vu en classe de première.

1 bit permet de stocker soit un 0 soit un 1  
1 octet est un ensemble de 8 bits  
 $1\text{ Ko} = 1\ 000\ \text{octets}$   
 $1\text{ Mo} = 1\ 000\ \text{Ko} = 1\ 000\ 000\ \text{octets}$   
 $1\text{ Go} = 1\ 000\ \text{Mo} = 1\ 000\ 000\ \text{Ko}$   
 $= 1\ 000\ 000\ 000\ \text{octets}$

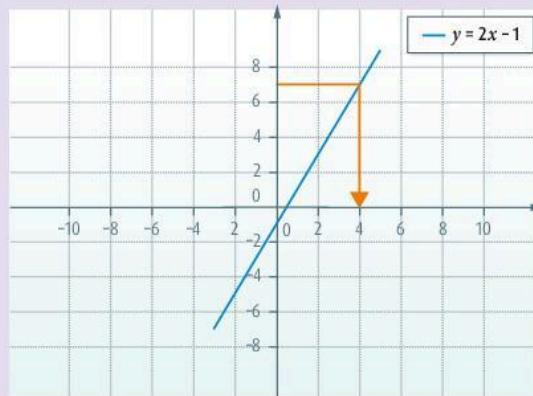
### Maths 2<sup>de</sup> Fonctions affines

Une fonction permet d'exprimer la dépendance d'une variable par rapport à une autre. Dans le cas d'une fonction affine,  $f(x) = mx + p$ , si on trace la courbe représentative de  $y = f(x)$ , on peut :

1. déterminer graphiquement  $y$  connaissant  $x$



2. déterminer graphiquement  $x$  connaissant  $y$



3. – déterminer  $y$  connaissant  $x$  par le calcul ou à l'aide d'un langage de programmation

$$x = 5 \text{ et } y = f(x) = 2x - 1 \text{ donc } x = 2 \times 5 - 1 = 9$$

– déterminer  $x$  connaissant  $y$  par le calcul ou à l'aide d'un langage de programmation

$$y = 3 \text{ et } y = f(x) = 2x - 1 \text{ donc } 3 = 2x - 1 \text{ d'où } 2x = 4 \text{ et donc } x = 2$$

## Se tester avant de démarrer

### Savez-vous répondre aux questions suivantes ?

1. Peut-on reproduire fidèlement un signal analogique (comme un son) lors d'une numérisation ?
2. La compression d'un fichier s'accompagne-t-elle forcément de pertes d'information ?
3. Avez-vous déjà utilisé un algorithme d'intelligence artificielle ? Si oui dans quel but ?

**Des robots champions du monde.** Ces robots, développés par des étudiants en informatique de l'université de Leipzig en Allemagne, ont remporté la coupe du monde "RoboCup 2018" à Montréal, après un match serré mené 1-0 contre une autre équipe allemande de l'université de Brême.





Comment à partir du traitement de l'information numérique a-t-il été possible de développer une intelligence artificielle ?

# Une brève histoire de l'informatique

L'informatique est aujourd'hui omniprésente à la fois avec des appareils aux tâches très spécifiques et des machines universelles dont les applications ne sont limitées que par leur vitesse de calcul.

Comment a évolué l'informatique de la naissance du concept jusqu'à aujourd'hui ?



## Vidéos du métier Jacquard



1 Le principe du tissage est d'entrelacer des fils afin d'obtenir une surface. Avant le métier à tisser Jacquard, on obtenait un tissu uni. Les motifs étaient ensuite brodés à la main.



2 Avec le métier Jacquard, on dessine d'abord le motif désiré sur papier quadrillé.



3 Chaque ligne du dessin est codée sur une carte perforée.

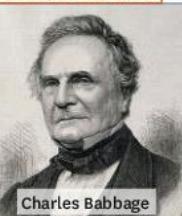


4 Lors du tissage, la carte perforée permet de sélectionner automatiquement les fils à relever afin de directement tisser le motif.

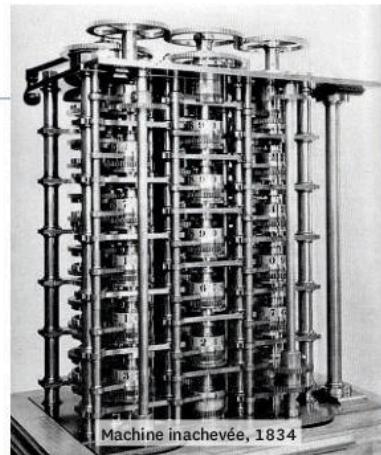
## DOC 1 Invention du métier Jacquard en 1801.



## Ada Lovelace, la première codeuse



À une époque où tous les calculs en science et technologie reposent sur des tables de valeurs, le mathématicien anglais Charles Babbage conçoit en 1834 une machine mécanique permettant d'automatiser un certain nombre de calculs, évitant ainsi les erreurs. Il fabrique un premier modèle de démonstration inachevé (photo). La machine analytique complète construite selon ses plans ne verra le jour qu'en 1991. Cette calculatrice universelle, programmable à l'aide de cartes perforées, permet de stocker les résultats des calculs pour les réutiliser. Ada Lovelace travaille avec Charles Babbage et publie en 1842 le premier programme informatique prévu pour s'exécuter sur cette machine.



## DOC 2 La machine de Babbage et Lovelace.

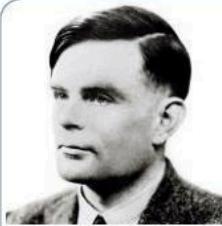
### Vidéo calculateur mécanique et mesure de temps de calculs



L'ENIAC conçu en 1943 par Presper Eckert et John Mauchly à l'université de Pennsylvanie, permet de calculer une trajectoire d'une table de tir presque instantanément grâce à l'automatisation des calculs. Les données sont introduites grâce à des cartes perforées, la programmation se fait grâce à la connexion de fils comme dans un standard téléphonique (voir photo) : c'est un travail long, fastidieux et difficile à vérifier. La puissance de calcul des ordinateurs n'a depuis cessé d'augmenter. Pour comparer, ces premiers ordinateurs de plusieurs tonnes n'ont même pas la puissance de calcul d'une calculatrice d'aujourd'hui.

Moyens employés	Durée d'une multiplication de 2 nombres de 10 chiffres	Temps de calcul d'une trajectoire d'une table de tir
Homme à la main, ou calculateur mécanique type machine de Babbage	5 min	2,6 j
Homme avec calculateur de bureau	10 à 15 s	12 h
ENIAC (électronique)	0,001 s	3 s
Unité de calcul d'un ordinateur des années 2000	30 ns	36 µs

## DOC 3 L'ENIAC, un calculateur automatique.



Le mathématicien anglais Alan Turing est une des figures les plus marquantes de l'histoire de l'informatique. En 1936, à 24 ans, il redéfinit de manière originale la notion de calcul: ce qui est calculable peut être décomposé en un nombre fini d'étapes, pouvant chacune être réalisée par une machine. Il pose également le concept d'une machine universelle capable de simuler n'importe laquelle de ces machines: la configuration nécessaire fait alors partie des données d'entrée. Cette machine, dite de Turing, est

capable de traiter n'importe quel type de données et constitue l'ancêtre des ordinateurs.

En 1942, grâce à la construction d'un calculateur, il décrypte le code Enigma utilisé par les nazis et joue ainsi un rôle crucial dans la guerre. Son texte «L'ordinateur et l'intelligence» publié en 1950 dans lequel il évoque le concept des machines pensantes, est considéré comme la base de l'intelligence artificielle. Le destin de ce génie est cependant brisé en 1952 lors de son procès pour homosexualité. Condamné à la castration chimique, Turing se suicide deux ans plus tard en croquant une pomme imbibée de cyanure.

#### DOC 4 Alan Turing et la machine universelle.



En 1945, le mathématicien Von Neumann pose le principe d'une architecture novatrice des calculateurs. Le processeur (élément réalisant le traitement des données) est divisé en deux parties: l'unité de commande, qui organise le séquencement des instructions, et l'unité arithmétique, qui exécute ces instructions. De plus, les instructions, au lieu d'être codées sur un support externe (ruban, carte, tableau de connexions), sont enregistrées dans la mémoire comme le sont les données. Ainsi, un programme enregistré peut être traité comme une donnée par d'autres programmes. Cette idée, esquissée dans la machine de Turing, trouve ici sa concrétisation. L'architecture de Von Neumann est encore utilisée aujourd'hui malgré quelques modifications mineures, notamment parce que les processeurs modernes comportent plusieurs coeurs (plusieurs processeurs sur une même puce).

**DOC 5 L'architecture machine de Von Neumann.** On retrouve dans cette architecture les différentes unités nécessaires à un système de traitement de l'information: une unité d'entrée, une unité de stockage (mémoire), une unité de traitement et une unité de sortie.

Objet	Programmable par l'utilisateur	Données
GPS	Non	<b>Entrées:</b> signaux GPS <b>Sorties:</b> données de positionnement et cartes
Console de jeux	Non	<b>Entrées:</b> appui sur les touches, programme de jeu <b>Sorties:</b> images, sons
Calculatrice	Oui	<b>Entrées:</b> fichiers, valeurs chiffrées, appui sur les touches <b>Sorties:</b> courbes, valeurs chiffrées
Ordinateur	Oui	<b>Entrées:</b> texte, son, image, vidéo, entrée clavier, souris, etc. <b>Sorties:</b> texte, son, image, vidéo, etc.
Smartphones	Oui	<b>Entrées et sorties</b> variées: texte, son, image, vidéo, positionnement, données des capteurs, etc.

#### DOC 6 Le multimédia et l'informatique embarquée.

On parle d'informatique embarquée dans tout appareil dont le fonctionnement repose, au moins en partie, sur un ou des programmes informatiques.

### EXPLOITER LES DOCUMENTS

- À l'aide d'un logiciel graphique ou de présentation (type Powerpoint), placez sur une frise les grands événements de l'histoire de l'informatique présentés dans cette double page.
- Rédigez un paragraphe expliquant les apports d'Ada Lovelace, d'Alan Turing et de John Von Neumann à l'informatique moderne.
- Indiquez comment les capacités des machines traitant de l'information ont évolué au cours de l'histoire. Vous raisonnerez sur la rapidité et la nature de l'information traitée.
- Proposez trois autres appareils de la vie courante dans lesquels de l'informatique est utilisée. Vous préciserez si l'appareil est programmé par le concepteur ou l'utilisateur (DOC. 6).

### ESPRIT CRITIQUE

**Le rôle des femmes dans les découvertes informatiques a longtemps été passé sous silence.**

→ Présentez trois courts portraits de femmes qui, comme Ada Lovelace, ont été à l'origine d'innovations marquantes dans ce domaine.

Pistes de travail ► Recherche Internet

## Programmes et données

Les ordinateurs multimédia traitent aujourd’hui un nombre très important de données grâce à des programmes. Sur les appareils mobiles, le problème de stockage se pose souvent car ces données sont de plus en plus nombreuses et volumineuses.

**Quelle est la différence entre donnée et programme ?**  
**Quel est l'espace mémoire utilisé par les différents types de données ?**

Type	Extension	Taille	Exemples d'exécutables utilisables
Son	WAV, MP3, FLAC	Quelques Mo pour un morceau de musique de quelques minutes	iTunes, Windows media player
Image	JPG, BMP, PNG, GIF	Quelques centaines de Ko à plusieurs Mo	Gimp, Apple Photos, Picasa, Paint, Adobe photoshop
Vidéo	AVI, MOV, MP4	La taille monte très vite en fonction de la durée. De quelques dizaines de Mo jusqu'à plusieurs Go	Windows media player, Quicktime, VLC

**DOC 1 Diversité et taille des données multimédia.** Le premier type d'informations que les systèmes étaient capables de manipuler étaient des nombres, puis des textes, jusqu'à l'arrivée dans les années 1980 des systèmes multimédia, c'est-à-dire capables de manipuler divers types d'informations : images, sons, vidéos... Les instructions d'un programme informatique sont également des informations et sont stockées dans le même type de mémoire que les données.

Tout fichier informatique est stocké sous forme binaire, c'est-à-dire sous la forme d'une suite de 0 et de 1. Pour un fichier texte, il faut donc utiliser une table de correspondance entre chaque caractère et un code binaire. Historiquement, la première table appelée ASCII permettait de coder :

- ▶ les lettres minuscules et majuscules de l'alphabet latin non accentuées;
- ▶ les 10 chiffres;
- ▶ quelques caractères de ponctuation;
- ▶ une série de caractères qui ne s'affichent pas, comme le retour à la ligne par exemple.

Aujourd'hui, on utilise plutôt des encodages en UTF-8 ou en ISO-8859-15 qui codent plus de caractères, qui sont compatibles avec la table historique mais pas forcément entre eux.

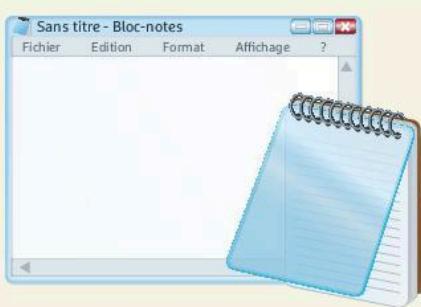
Caractère	ASCII	ISO-8859-15	UTF-8
A	1 octet: 01100001	1 octet: 01100001	1 octet: 01100001
é	Absent de la table	1 octet: 1101001	2 octets: 11000011 10101001

**Animation : affichage de la table ASCII**



**DOC 3 Jeu de caractères encodés sur 8 bits dans la norme ISO-8859-15.** Cette table regroupe les caractères de l'alphabet latin, ainsi que les 10 chiffres des lettres accentuées et des signes de ponctuation. Elle recouvre ainsi les caractères nécessaires à l'écriture des langues d'Europe occidentale.

**DOC 2 Différents encodages des caractères.**



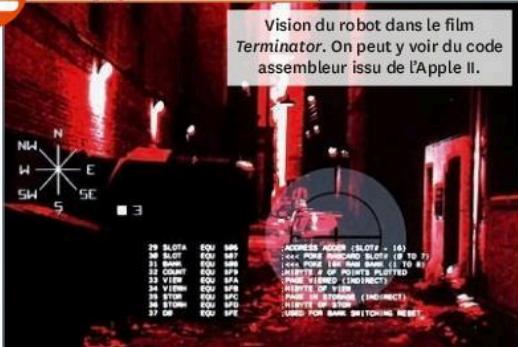
### PROTOCOLE

- Ouvrir BlocNote
- Incrire dix fois la lettre A sous la forme AAAAAAAA
- Enregistrer le fichier sous le nom donnees.txt
- Fermer BlocNote
- Clic droit sur le fichier ⇒ Propriétés : cela donne accès à la taille du fichier.
- Essayer avec une autre lettre à la place du A, et avec un mélange de lettres.

**DOC 4 Construire un fichier de données sous Windows.**

**Liste d'extensions à risque**

À partir d'un problème à résoudre par l'informatique, la première étape est de décomposer ce problème en une suite d'opérations simples décrites en français. C'est l'**algorithme**. Cet algorithme peut ensuite être traduit dans un langage de programmation compréhensible par le programmeur comme Scratch ou Python pour obtenir un **programme**. Le **fichier exécutable** est le fichier qui contient le programme sous une forme compréhensible par l'ordinateur (langage machine) ou interprétable directement par l'ordinateur, c'est à dire traduit en langage machine à la volée au fur et à mesure de l'exécution (programme Python avec le bon interpréteur ou bytecode Java par exemple). Le langage machine est constitué d'instructions très simples spécifiques au processeur et compréhensibles directement par l'unité de commande du processeur. Les **données** sont les informations fournies au programme pour qu'il puisse travailler. Un fichier exécutable est une donnée pour un autre programme : le système d'exploitation.

**DOC 5** Quelques définitions utiles.**Le langage machine en vidéo**

Vision du robot dans le film *Terminator*. On peut y voir du code assembleur issu de l'Apple II.

(mémoire locale), comparer deux valeurs, les ajouter, etc. Ces instructions sont codées en binaire : c'est le langage machine. Seul le langage machine est compréhensible directement par le processeur, mais il est incompréhensible pour le programmeur, c'est pourquoi il peut aussi être codé sous forme de mots : c'est le langage Assembleur. Ce langage n'est pas naturel pour un humain et nécessite beaucoup de lignes de code pour exécuter des opérations relativement simples, on va donc utiliser en pratique plutôt des langages de haut niveau comme Python ou Scratch. Les programmes écrits en langage de haut niveau seront dans tous les cas convertis en langage machine, avant ou pendant l'exécution par l'ordinateur.

**DOC 7** Qu'est-ce que le langage Assembleur ?**EXPLOITER LES DOCUMENTS**

- Donnez la taille du fichier **donnees.txt** et indiquez si cette taille change en fonction des lettres (**DOC. 4**).
- Les fichiers **donnees.txt** et **executable.bat** ont été construits avec un même logiciel. Expliquez les différences entre ces deux fichiers.
- Une page de texte contient environ 2 500 signes. En ne considérant que des caractères présents dans la table ASCII, indiquez la taille en mémoire du fichier considéré (**DOCS 2 à 4**).
- Justifiez la taille du fichier **soutie.txt** (**DOC. 8**).
- Pour stocker un livre au format numérique, expliquez s'il vaut mieux filmer les pages, les scanner ou les avoir dans un format texte.

Voici une liste d'extensions de fichiers exécutables avec les systèmes d'exploitation correspondants. Les plus courants sous Windows sont .exe et .msi mais il en existe d'autres comme le .bat par exemple. Ces fichiers sont exécutables et contiennent donc des opérations réalisées par le système d'exploitation. Si, en général, ces fichiers réalisent le programme souhaité, ils peuvent être parfois détournés pour réaliser des actions malveillantes. Il existe d'autres extensions à risque, et certains fichiers textes ou feuilles de calcul peuvent aussi contenir du code exécutable.

Extension	Système d'exploitation
.apk	Android
.app	MacOS
.bat	Windows
.bin	Windows, macOS, Linux
.exe	Windows
.ipa	iOS
.sh	MacOS, Linux

**DOC 6** Quelques extensions d'exécutables.**PROTOCOLE**

- Ouvrir BlocNote
- Inscrire les deux lignes suivantes:  
 set /p lettre=Entrez une lettre:  
 for /l %%i in (1, 1, 10) do (echo %letter%>>soutie.txt)
- Enregistrer sous : modifier le type (afficher « Tous les types »)
- Donner le nom **executable.bat** et cliquer sur Enregistrer
- Le fichier obtenu est exécutable, double cliquer dessus pour vous en rendre compte.
- Entrer une lettre. L'exécutable fabrique un fichier **soutie.txt** qui contient dix fois la lettre choisie (une par ligne).
- Noter la taille du fichier **soutie.txt**.

**DOC 8** Construire un fichier exécutable

**sous Windows.** Voici un protocole pour écrire un programme dans le langage Batch de Windows. Ce programme est exécutable directement sur les machines Windows grâce à un interpréteur présent dans le système d'exploitation : cmd.exe.

**ESPRIT CRITIQUE**

Pour mettre à l'abri leurs données (contacts, applications, musiques, photos...) de nombreux utilisateurs de smartphone choisissent un stockage sur des *cloud*.

→ Où sont physiquement stockées les données des *cloud*?

→ Le stockage de données en ligne est-il neutre d'un point de vue écologique?

**Pistes de travail** ► **Recherche Internet**

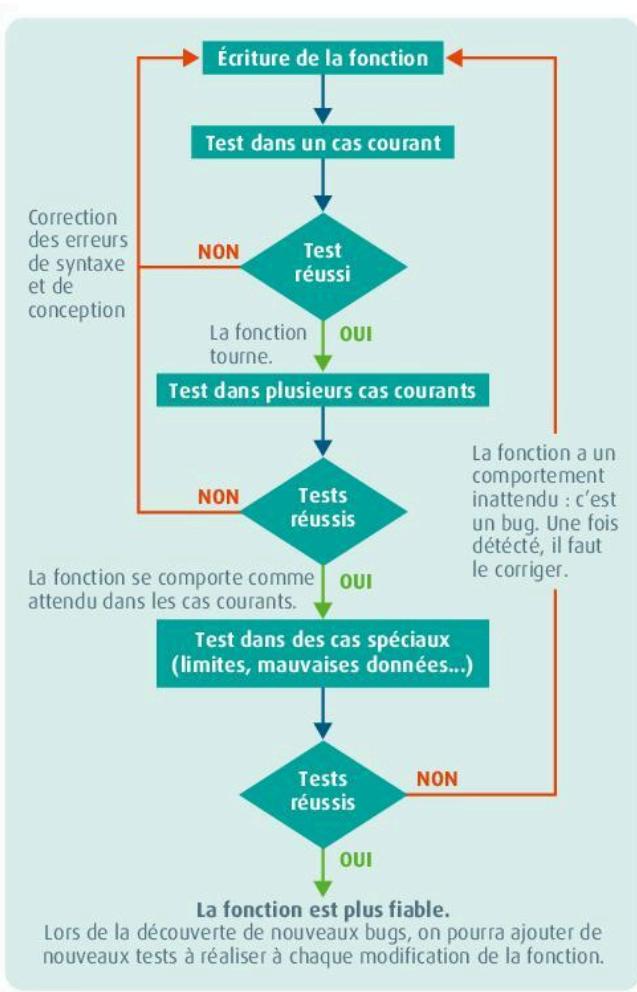
## Il y a un bug !

Même si les pratiques ont évolué pour produire du code de qualité, il est impossible de garantir un projet informatique ne comportant aucun bug. Ceux-ci peuvent causer des problèmes allant du petit désagrément aux accidents humains graves.

**Quels sont les impacts sociétaux des bugs informatiques et comment peut-on les corriger ?**

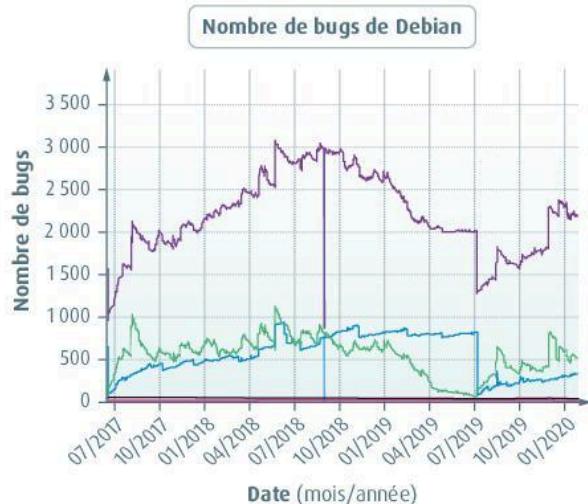


**DOC 1 Le cas de Toyota.** Toyota a dû rappeler au garage 1,9 millions de véhicules en 2014 et 625 000 de plus en 2015 pour une mise à jour du logiciel d'une trentaine de minutes. Cette opération très coûteuse pour le constructeur a été rendue nécessaire par la présence d'un bug qui entraînait l'arrêt de la voiture dans certaines conditions.



**DOC 2 Identifier les bugs d'un programme.**

Certains projets comme l'écriture d'un système d'exploitation nécessitent plusieurs millions de lignes de code. Le dernier noyau Linux stable compte par exemple environ 20 millions de lignes de code. Un projet de cette ampleur comporte nécessairement des bugs découverts et corrigés au fur et à mesure. Ils sont classés puis corrigés en fonction de leur严重性和出现频率. On peut suivre à l'adresse <https://bugs.debian.org/release-critical/> le nombre de bugs critiques en cours de résolution pour une distribution Linux appelée Debian. Au bout d'un certain temps, et après beaucoup d'efforts, on obtient du code très fiable. Par exemple pour la version précédente (courbe rose), il ne subsiste environ qu'un bug pour chaque million de lignes de code.



- Nombre de bugs concernant la version précédente : 28
- Nombre de bugs pour le moment ignorés : 64
- Nombre total de bugs critiques dans toutes les versions disponibles : 2 185
- Nombre de bugs dans la version courante de Debian : 326
- Nombre de bugs concernant la prochaine version : 525

**DOC 3 Présence de bugs.**



### Ajouter des tests et corriger une fonction

Voici un programme Python permettant de calculer la moyenne des éléments d'une liste.

```
def moyenne(liste) :
    somme = 0
    for val in liste :
        somme = somme + val
    return somme / len(liste)
```

# On calcule la somme  
# de tous les éléments de la liste  
# en les ajoutant un par un  
# On renvoie la moyenne, correspondant à cette somme  
# divisée par le nombre d'éléments de la liste

Nous allons appliquer les tests suivants qui doivent vérifier tous les comportements de cette fonction.

On va donc vérifier que moyenne([1,2,3]) renvoie bien 2.0

moyenne([1,2,3]) #Renvoie 2.0 donc la fonction se comporte correctement

Faisons maintenant le test avec une liste vide: moyenne([])

On obtient le message d'erreur assez clair suivant:

line 5, in moyenne return somme / len(liste) ZeroDivisionError: division by zero

**Solution:** Il faut ajouter une condition dans le programme car, dans le cas où la liste ne contient pas d'éléments, on ne peut pas diviser par 0.

On ajoute donc les trois lignes ci-dessous au début de la fonction pour la corriger.

Puis on teste la fonction avec cette nouvelle définition:

moyenne([]) # Renvoie maintenant None et plus une erreur

```
if len(liste) == 0 :    # Si la liste est vide
    return None           # Alors on renvoie None
else :                  # Sinon
```

### DOC 4 Exemple de tests en Python.

La fonction moyenne précédente n'est pas encore parfaite. Par exemple si, par erreur, on lui envoie une liste de lettres, elle renverra un message d'erreur.

On peut la corriger comme suit:

### DOC 5 Correction de la fonction moyenne.

```
def moyenne(liste) :
    if len(liste) == 0 :    # Cas d'une liste vide
        return None          # renvoie None
    else :
        somme = 0
        for val in liste :
            if type(val) == int or type(val) == float : # Cas d'un nombre dans la liste
                somme = somme + val                      # On calcule la moyenne
            else :                                     # Cas d'un autre type de donnée dans la liste
                return None                         # On renvoie None car on ne peut pas calculer de moyenne
        return somme / len(liste) # Si le calcul est mené à son terme, on renvoie la moyenne
```

Voici deux exemples de fonctions qui contiennent des bugs.

def val\_frac(a, b) :  
 return a / b

```
def sqrt(val) :
    return val ** 0.5    # soit val à la puissance 0.5
                        # c'est-à-dire sa racine carrée
```

### DOC 6 Des bugs à identifier.

## EXPLOITER LES DOCUMENTS

- Indiquez si Toyota peut certifier que ses logiciels ne contiennent plus de bugs ([DOC. 1](#)).
- Avec l'essor de l'informatique embarquée, les autres constructeurs sont-ils à l'abri de bugs de ce type ? Discutez de l'impact sociétal d'un bug dans le cas d'un véhicule autonome.
- Proposez des couples a, b et des valeurs de val permettant de tester les fonctions proposées dans le [DOC. 6](#).
- Expliquez dans un court paragraphe les problèmes posés par la fonction initiale du [DOC. 4](#) et comment ils ont été corrigés ([DOCS 4 et 5](#)).
- Déterminez les situations qui pourraient poser problème pour les fonctions du [DOC. 6](#). Proposez alors des modifications à apporter afin d'éviter les bugs.

## ESPRIT CRITIQUE

Des bugs célèbres ont eu lieu dans l'[histoire des avancées scientifiques et technologiques](#).

→ Quel a été l'impact humain et économique des deux bugs célèbres ci-dessous ?

- celui survenu sur un vol de la fusée Ariane 5 ;
- celui de l'ordinateur de bord de véhicules de la marque Lexus.

[Pistes de travail](#) ► [Recherche Internet](#)

# Que puis-je faire pour vous ?

L'intelligence artificielle (IA) nous entoure, de manière visible (smartphones, assistants personnels, ...) mais aussi comme une partie de nombreux logiciels. L'apprentissage machine, un des domaines de l'IA, utilise en masse des données disponibles pour produire des modèles prédictifs.

**Qu'est-ce que l'intelligence artificielle et comment peut-on l'utiliser ?**



**Interview de Serge Abiteboul,  
directeur de recherche à l'INRIA**

L'IA regroupe des techniques informatiques permettant à des machines de réaliser des actions normalement associées à l'intelligence humaine : compréhension, adaptation, apprentissage... Une branche de l'IA, l'apprentissage machine, a récemment obtenu des résultats surprenants pour des problèmes jusqu'ici sans solution informatique. Un programme d'apprentissage est un programme particulier (ex: un «réseau de neurones») avec de très nombreux paramètres à régler. On lui donne comme entrée une image de chat et on lui dit qu'il faut régler ses paramètres pour que la sortie

dise : c'est un chat. On fait la même chose avec plein d'images de chats et de chiens. À l'aide de tas de calculs, le programme règle ses paramètres pour reconnaître les chats et les chiens sur ces images. Quand on lui donne l'image d'un animal inconnu, le programme détermine si ce nouvel animal ressemble plutôt à un chat ou un chien : il a appris à distinguer les deux espèces. Cela marche pour d'autres problèmes. Il faut bien saisir l'importance de toutes ces images qui ont permis au programme d'apprendre. On appelle cela une «supervision». L'apprentissage non supervisé s'appuie sur d'autres techniques, comme l'apprentissage par renforcement par lequel le programme apprend de ses succès et de ses erreurs.

## DOC 1 Intelligence artificielle et apprentissage machine.

**Teachable machine de Google et protocole**



Les voitures autonomes possèdent des algorithmes qui fonctionnent sur des ordinateurs de bord puissants. Ces algorithmes, utilisant des réseaux neuronaux artificiels, permettent à la machine d'apprendre à reconnaître des objets sur des images. Ainsi, à partir d'un nombre important de données accumulées par les caméras et différents capteurs, la voiture autonome apprend à reconnaître des véhicules, des piétons, des panneaux de signalisation etc... La teachable machine de Google fonctionne comme un réseau neuronal présent dans les voitures autonomes. La caméra de l'ordinateur ou du smartphone

est utilisée pour l'apprentissage (grande quantité d'images obtenues en très peu de temps).

L'expérience a ici été réalisée en utilisant des jouets :

- une voiture de pompier ;
- un personnage ;
- un animal.

Après quelques secondes d'acquisition sur chaque objet sous différents angles, le réseau de neurones a accumulé assez de données pour reconnaître un véhicule, un animal et un piéton autres que ceux utilisés pour l'apprentissage.

### 1 Données d'entrée

### 2 Apprentissage

Pour classer les futures entrées :

- piéton,
- animal,
- véhicule

#### Training

Model Trained

Retardé

### 3 Tests

Deux exemples de sorties.

Output	Piéton	Animal	Véhicule
Échec des tests	~40%	~60%	~10%
Tests réussis	~95%	~5%	~0%

### 4 Utilisation du modèle

Utilisation du modèle

## DOC 2 Entrainer une IA.

**Jeu de données et animation**

Parce qu'on dénombre plus de 30 universités dans la ville et ses environs, Boston attire chaque année des étudiants du monde entier. Un étudiant souhaite s'installer dans un quartier de

Boston pour lequel il n'a pas de données sur le prix des logements. Comment l'intelligence artificielle peut-elle l'aider à estimer le prix d'un logement?

### 1 Collecte d'un jeu de données

On utilise ici un jeu de données sur les logements dans la banlieue de Boston, dans lequel les prix et différents paramètres d'entrée sont connus.

**But:** connaissant les paramètres d'entrée d'une zone de prix inconnu, estimer le prix des logements dans cette zone.

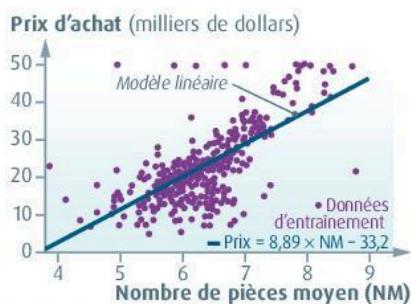
Données	Paramètres d'entrée					Prix (k\$)
	Criminalité (U.A.)	Pollution (U.A.)	Nombre de pièces moyen	...		
1	0,00632	0,538	6,575	...	24,0	
2	0,02731	0,469	6,421	...	21,6	
3	0,02729	0,469	7,185	...	34,7	

### 2 Entraînement (sur 90 % des données)

- Séparation des données: 90 % pour l'entraînement et 10 % pour le test.
- Représentation du prix en fonction d'un paramètre.

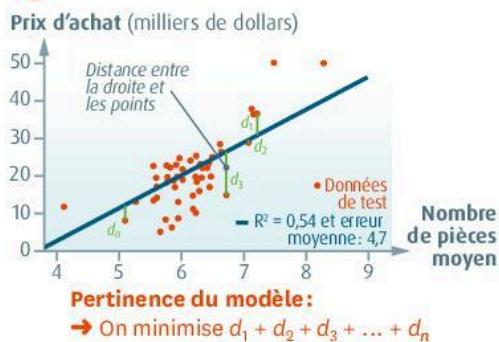
#### Pertinence du modèle:

- Évaluation visuelle (reflet des données)
- Critères quantifiés (coefficient de détermination  $R^2$  proche de 1 lorsque le modèle représente bien les points)

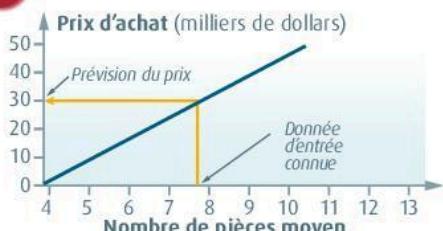


Répétition des étapes 2 à 3 jusqu'à l'obtention d'un modèle pertinent

### 3 Test (sur 10 % des données)

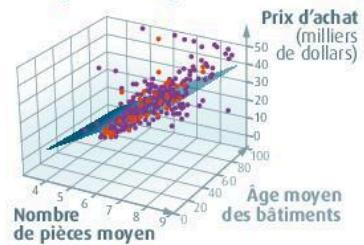


### 4 Utilisation du modèle



### 5 Utilisation de plusieurs paramètres

- Étapes 2 à 4 avec plusieurs paramètres
- Amélioration du modèle



### DOC 3 Le fonctionnement de l'apprentissage machine.

## EXPLOITER LES DOCUMENTS

- Expliquez dans un court paragraphe les points communs de la démarche du **DOC. 2** et du **DOC. 3**.
- Expliquez en quoi le travail effectué sur les données dans les **DOCS 2 et 3** relèvent de l'IA (**DOC. 1**).
- Indiquez la corrélation exploitée dans le jeu de données du **DOC. 3**. Expliquez si, selon vous, ce modèle pourrait encore être amélioré et si oui, comment.
- Indiquez si le modèle proposé dans la vignette **5** pourrait être encore amélioré et si oui, comment (**DOC. 3**).
- En utilisant le modèle des vignettes **2 à 4** du **DOC. 3**, estimez, en utilisant l'équation du modèle, la valeur attendue pour le prix d'un logement à Boston pour un nombre moyen de pièces de 6,75. Même question par lecture graphique.

## ESPRIT CRITIQUE

L'apprentissage machine utilise un nombre important de données pour entraîner les modèles.

- D'où proviennent les données exploitées dans le **DOC. 3**? Où sont-elles stockées et qui en sont les créateurs ?
- D'où peuvent provenir à votre avis les données utilisées pour entraîner une IA ?

**Pistes de travail** ► **Lien numérique du DOC. 3 et UNITÉ 6 p. 294**

# Le véhicule autonome vraiment intelligent ?



## 1. Classification des véhicules autonomes.

La NHTSA (National Highway Traffic Safety Administration) américaine classe les véhicules en 6 catégories en fonction de leur niveau d'autonomie. Ce même classement est utilisé par les constructeurs automobiles et la réglementation européenne. Les aides à la conduite disponibles pour la sécurité ou le confort (régulateur de vitesse adaptatif par exemple) ont fait progressivement augmenter le niveau d'autonomie des véhicules. La législation évolue pour permettre l'arrivée des premiers véhicules complètement autonomes pour 2025.

Niveaux de conduite autonome						
Conduite autonome ↔ Conduite manuelle	D'après www.nhtsa.gov					
	Niveau 0 Conducteur seul	Niveau 1 Conducteur assisté	Niveau 2 Autonomie partielle	Niveau 3 Autonomie conditionnelle	Niveau 4 Haut niveau d'autonomie	Niveau 5 Conduite complètement autonome
Pas d'intervention active du véhicule.	L'autre mouvement est assuré par le véhicule.	Le véhicule assure le mouvement longitudinal et latéral dans certains cas bien définis.	Le véhicule assure le mouvement longitudinal et latéral dans certains cas bien définis. Le système peut demander au conducteur de récupérer le contrôle total avec une marge de temps suffisante.	Le conducteur n'a à surveiller tout le temps ni le mouvement ni l'environnement du véhicule mais doit toujours être en position de récupérer le contrôle total du véhicule.	Le conducteur n'est pas requis dans certains cas bien définis.	Le véhicule assure le mouvement longitudinal et latéral dans toutes les situations sans besoin de la présence d'un conducteur.

## 2. Conduite autonome avancée chez Tesla.



Sur son site, le constructeur Tesla présente les avancées dans le développement de son modèle de voiture autonome : le système est composé de huit caméras panoramiques qui offrent une visibilité à 360° autour de la voiture avec une portée de 250 m. À cela viennent s'ajouter douze capteurs à ultrasons nouvelle génération permettant de détecter les objets solides ou les liquides avec une portée et une précision quasiment deux fois supérieure au système précédent. Un radar avant capable de voir à travers de fortes pluies, le brouillard, la poussière et même la voiture qui précède complète les données environnementales. Toutes ces données sont traitées par une IA présente dans l'ordinateur de bord du véhicule qui prend alors les décisions et agit sur les commandes du véhicule.

D'après [www.tesla.com](http://www.tesla.com)

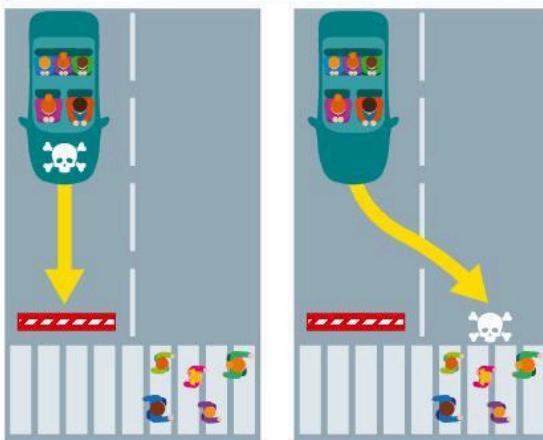


Conduite autonome en vidéo

**«La majorité des accidents de la route étant imputable à des erreurs humaines, le développement des voitures autonomes vise à réduire jusqu'à 90 % le nombre d'accidents.»** D'après Bonnefon et al, *Science*, 2016.

### 3. Le dilemme éthique de la voiture autonome.

Qu'est ce que la voiture autonome devrait faire ?



La majorité des accidents de la route étant imputable à des erreurs humaines, le développement des voitures autonomes vise à réduire jusqu'à 90 % le nombre d'accidents. Cependant, étant donné le nombre d'usagers de la voiture, le nombre d'accidents ne pourra pas être réduit à zéro. Cela soulève différents problèmes éthiques qui existaient bien avant la voiture autonome. Un obstacle non anticipé apparaît sur la route et il est trop tard pour le contourner. Que doit faire la voiture autonome ? Sauver à tout prix ses passagers, même au prix de la vie de piétons ? Ou sacrifier les passagers pour sauver les piétons ? Une telle situation est un dilemme éthique, c'est-à-dire une situation où les valeurs et les principes entrent en opposition et rendent les décisions difficiles.



Dossier de presse: Dilemme éthique de la voiture autonome

### 4. Du dilemme éthique au dilemme social.

Reprendons l'exemple du doc. 3. Dans une voiture classique, le conducteur agit instinctivement. Cette réaction de panique n'est pas interprétée comme une action prémeditée de mauvaise intention. En revanche, pour une voiture autonome, un dilemme éthique apparaît. Qui la voiture doit-elle sauver entre passagers et piétons ? La programmation en avance de la personne à sacrifier n'est-elle pas un homicide prémedité ? Qui sera responsable ? Le programmeur ? Le constructeur de la voiture ? L'utilisateur ?

Une étude publiée dans la revue *Science* en 2016 et faite par des chercheurs en France et aux USA montre que la grande majorité des 1 928 participants interrogés pense que la voiture autonome doit être programmée pour minimiser le nombre de victimes dans toutes les situations d'accident. En revanche, aucun participant ne souhaite acheter une voiture qui ne protégerait pas en priorité la vie de ses passagers. La législation des voitures fait donc face à un dilemme social : nous sommes d'accord pour minimiser les victimes à condition que les victimes soient les autres. Certains constructeurs, comme Mercedes-Benz en 2016, ont déjà tranché : en cas d'accident mortel, la voiture autonome privilégiera la protection de ses passagers.

### 5. Des conséquences économiques et sociétales positives ?



Rapport sur le développement de la voiture autonome

En mai 2018, le gouvernement publie un rapport sur le développement du véhicule automatisé. Il y est écrit que suite à la modification de la loi européenne en 2016, la loi française permettra la circulation de véhicules autonomes de niveau 4 d'ici 2022. Il est aussi indiqué que face à cette révolution technologique, l'action publique et les acteurs économiques cherchent à anticiper et créer les conditions pour que cette évolution soit, autant que possible, créatrice d'emplois. Les conséquences du développement des voitures autonomes pourraient concerner un grand nombre d'activités au moins dans les transports (700 000 emplois potentiels) et dans la filière industrielle automobile (500 000 emplois potentiels). Le développement de l'automatisation va probablement conduire à des transferts de qualifications, avec des opportunités de créer de nouveaux emplois et de nouvelles compétences à l'interface des transports et du numérique.

#### AIDE POUR RÉUSSIR LA MISSION

- Définissez la voiture autonome et comment l'IA permet de développer ces voitures (docs 1 et 2).
- Déterminez les dilemmes soulevés (docs 3 et 4).
- Indiquez les impacts économiques potentiels de cette technologie (doc. 5).



Le dilemme éthique de la voiture autonome en vidéos

# IA : vers une utilisation éthique des données

**Les faits :** Des millions de données personnelles sont utilisées pour l'intelligence artificielle.

**VOTRE MISSION** Expliquez dans un exposé oral de 5 min les biais et les problèmes éthiques de l'apprentissage machine.

## 1. Qu'est-ce qu'une donnée personnelle ?

La CNIL, Commission Nationale Informatique et Libertés, est une autorité administrative indépendante créée en 1978 qui veille à la protection des données personnelles.

Elle définit une donnée personnelle ainsi : « Il s'agit de toute information relative à une personne physique identifiée ou identifiable, directement ou indirectement, par référence à un numéro d'identification (ex.: n° de sécurité sociale) ou à un ou plusieurs éléments qui lui sont propres (ex.: nom et prénom, date de naissance, éléments biométriques, empreinte digitale, ADN...) ».

D'après [www.cnil.fr](http://www.cnil.fr)

## 2. Comment sont protégées mes données personnelles ?

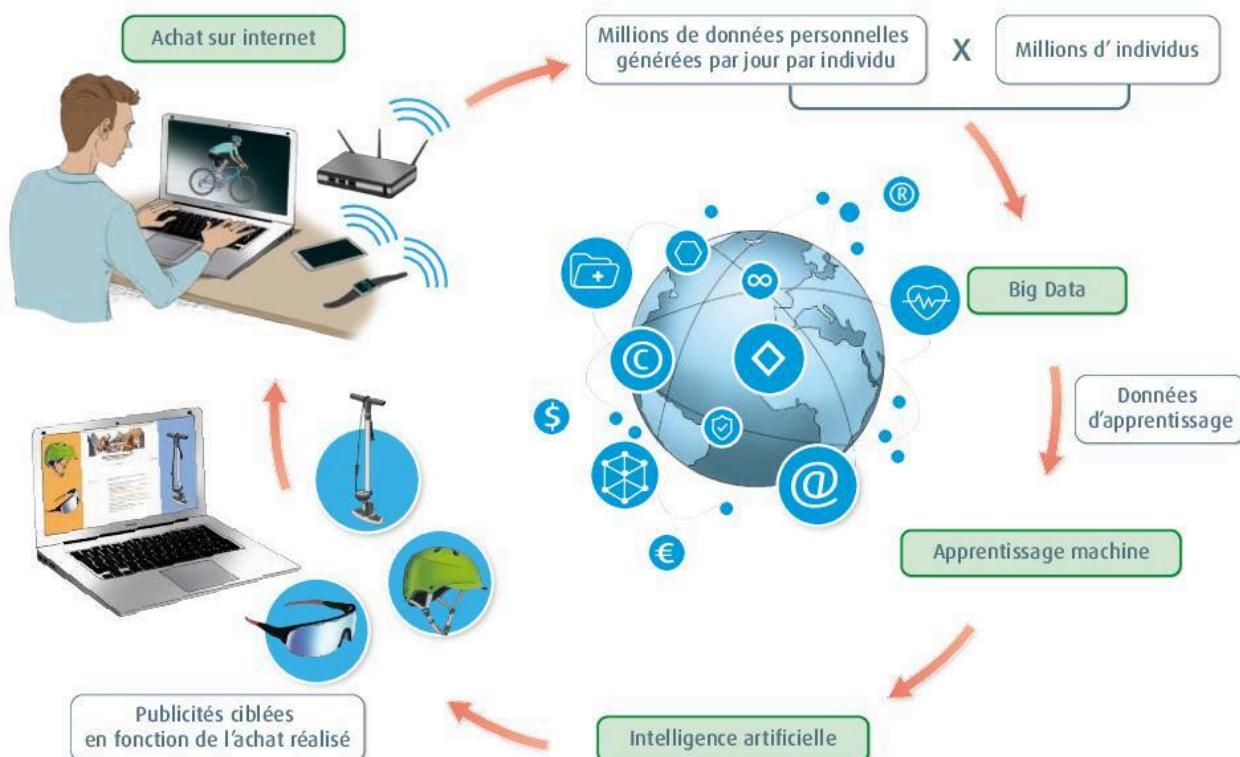
Le Règlement Général de l'Union européenne sur la Protection des Données, ou RGPD, est entré en vigueur le 25 mai 2018. Son objectif est d'harmoniser les lois sur la protection des données au sein de l'UE. Ce règlement stipule notamment que la prise de décision automatisée utilisée par de nombreux algorithmes d'apprentissage automatique, par exemple pour la publicité, n'est pas autorisée sans le **consentement explicite** de la personne concernée. Par ailleurs, certaines données personnelles ne peuvent pas être utilisées pour l'apprentissage machine comme les « données à caractère per-

sonnel qui révèlent l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques, l'appartenance syndicale, le traitement des données génétiques, des données biométriques (...). Enfin, il est important de savoir que le RGPD garantit des droits face à l'utilisation de ses données personnelles, comme le droit d'accès, le droit à l'oubli (déférer une page web liée à son identité), le droit à la portabilité (récupérer ses données dans un format lisible par une machine), le droit de rectification ou d'opposition.

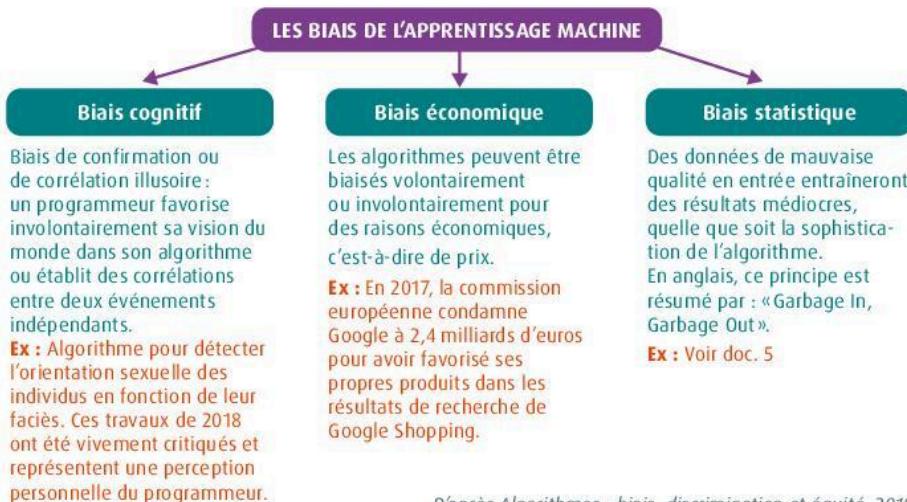
*Quels sont mes droits ?*



## 3. Big Data et IA.



## 4. Différents biais dans l'apprentissage machine.

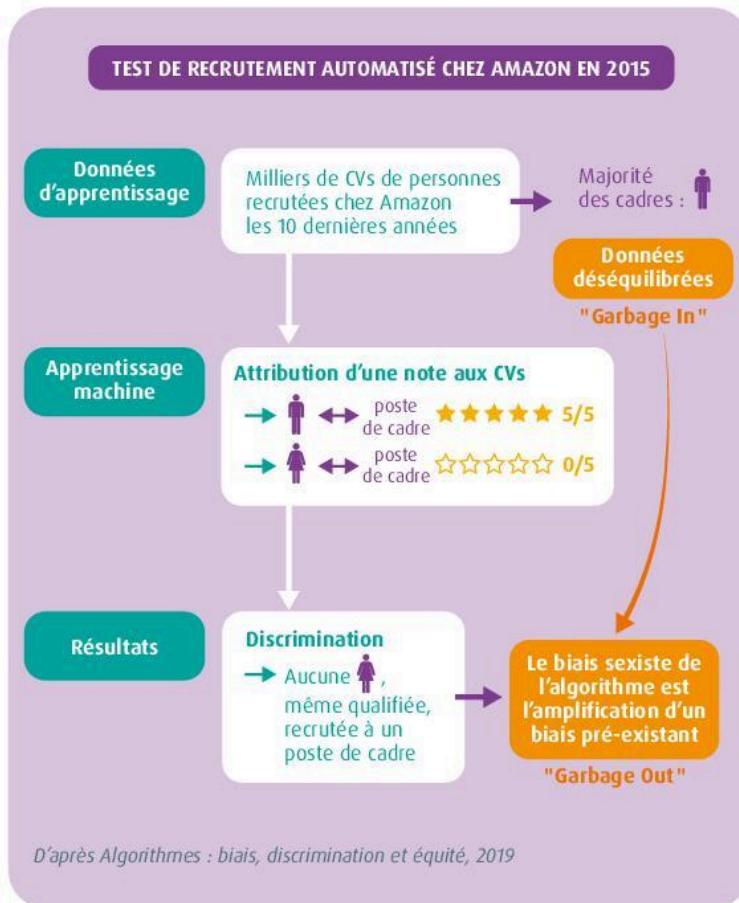


TIENS, LA RECONNAISSANCE FACIALE DE MON SMARTPHONE EST EN PANNE...



D'après Algorithmes : biais, discrimination et équité, 2019

## 5. Des biais statistiques.



D'après Algorithmes : biais, discrimination et équité, 2019

## 6. Les défis à relever.



Interview de Serge Abiteboul, directeur de recherche à l'INRIA

De plus en plus, des logiciels nous recommandent des produits à acheter, des contenus (vidéos, films...), nous aident à décider, voire décident pour nous. L'importance qu'ils acquièrent leur confère une responsabilité immense. Juridiquement, cela engage la responsabilité des humains et des entreprises qui conçoivent, déploient ces logiciels ou choisissent les données d'entraînement. Cela dit, nous sommes tous concernés. À nous d'exiger que la qualité des logiciels soit vérifiée. Un jeu vidéo qui plante, c'est désagréable, mais un algorithme qui amène à refuser la libération conditionnelle d'un condamné par erreur, ce n'est juste pas acceptable. Ça l'est encore moins si c'est parce que le logiciel ne fait qu'imiter les biais des données sur lesquelles il base sa recommandation. Quand les logiciels finissent par définir nos vies, l'absence de biais doit être garantie. Cela passe par la vérification (automatique) qu'ils satisfont les propriétés que nous jugeons indispensables, par des tests extensifs de ces propriétés. Les algorithmes permettent de transformer le monde. Notre responsabilité collective est de nous assurer que cela se fait pour le bien de tous.

### AIDE POUR RÉUSSIR LA MISSION

- Expliquez les règles d'utilisation des données personnelles (DOCS 1 à 3).
- Identifiez les différents types de biais de l'apprentissage machine (DOCS 4 et 5).
- Déterminez les défis à relever pour améliorer l'apprentissage machine (DOC. 6).

# L'inférence bayésienne

Vais-je obtenir une licence en trois ans, sachant que j'ai eu de bonnes notes au bac ? L'inférence bayésienne est une technique mathématique, fondée sur le théorème de Bayes, qui estime la plausibilité d'une hypothèse, étant donné des observations concrètes.



En théorie des probabilités, le théorème du mathématicien anglais Thomas Bayes permet de relier des probabilités dites conditionnelles : la probabilité d'une hypothèse étant donné des observations est reliée à la probabilité d'observations connaissant une hypothèse. L'inférence bayésienne est une technique de calcul qui repose sur le théorème de Bayes. Elle permet d'effectuer des estimations de probabilité au fur et à mesure que des observations nouvelles arrivent.

On connaît  $p(\text{«hypothèse»})$  sachant «**observation**». On déduit  $p(\text{«observation»})$  sachant «hypothèse».

**DOC 1** Thomas Bayes et son théorème.

**DOC 2** Détection de spams par inférence bayésienne.

Comment intervient l'inférence bayésienne dans l'apprentissage automatique ou les diagnostics médicaux ?

Arbre de probabilité



## 1. Apprentissage

- Tri manuel de spams par l'utilisateur
- Résultat indiqué à l'algorithme
- Calcul de probabilité que chaque mot soit associé à un spam par le programme

## 2. Inférence bayésienne

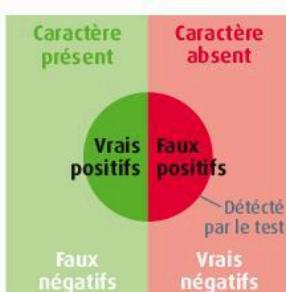
- Arrivée d'un nouveau courriel : l'algorithme calcule à l'aide de la formule de Bayes la probabilité que ce courriel soit un spam, sur la base de tous les mots qu'il contient.
- Si cette probabilité excède une valeur-seuil fixée à l'avance (par exemple 95 %), alors le courriel est classé en spam
- L'utilisateur peut indiquer qu'un courriel a été classé de manière erronée. Ainsi l'algorithme apprend davantage et peut s'améliorer.

Un courriel peut être classé comme spam alors qu'il ne l'est pas et inversement. Lorsqu'on applique une procédure de décision, nécessairement imparfaite, il apparaît des erreurs qu'on appelle des «faux», et qu'on qualifie de positif ou négatif selon le résultat de la détection. Les quatre possibilités se résument dans ce qu'on appelle un tableau de contingence.

Tableau de contingence

	Caractère présent	Caractère absent	*les faux parmi les positifs
Détection : test au résultat positif	Vrais positifs	Faux positifs*	Fréquence de faux positifs = $\frac{\text{Faux positifs}}{\text{Résultats positifs}}$
Pas de détection : test au résultat négatif	Faux négatifs	Vrais négatifs	Fréquence de faux négatifs = $\frac{\text{Faux négatifs}}{\text{Résultats négatifs}}$

Représentation sous forme de schéma



$$\text{Fréquence de faux positifs} = \frac{\text{Faux positifs}}{\text{Résultats positifs}}$$
  

$$\text{Fréquence de faux négatifs} = \frac{\text{Faux négatifs}}{\text{Résultats négatifs}}$$

	Spam	Courriels normaux
Détection de spam	932	60
Pas de détection de spam	419	2915

D'après Khan et Qamar, 2016.

**DOC 4** Un exemple de détection de spam. Après l'avoir entraîné, on applique l'algorithme à un ensemble de 4 326 courriels parmi lesquels 1 351 étaient du spam, et 2 975 normaux.

## DOC 3 Des faux, positifs et négatifs.

Les colonnes du tableau de contingence résument la présence du caractère (par exemple une maladie). Les lignes indiquent si le caractère est détecté ou non. On peut alors calculer le taux, ou fréquence, de faux positifs et de faux négatifs, qui caractérise la qualité de la procédure de décision. Les faux positifs et les faux négatifs sont définis comme les faux parmi les positifs et les négatifs, respectivement.

**Tableau de contingence**

		Caractère présent Ex : maladie, grossesse, etc.	Caractère absent Ex : individu sain
Détection : test au résultat positif	Vrai positif	Faux positif	
Pas de détection test au résultat négatif	Faux négatif	Vrai négatif	
Sensibilité =	Vrai positif	Spécificité =	Vrai négatif
Vrai positif + Faux négatif		Vrai négatif + Faux positif	

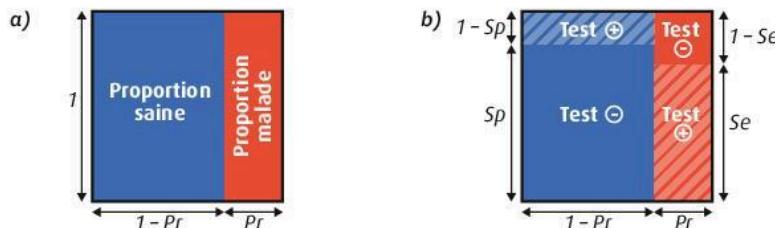
**En reprenant le schéma du DOC 3**

Sensibilité = 

Spécificité = 

**DOC 5 Sensibilité et spécificité d'un test diagnostique.**

La prévalence  $Pr$  est la proportion d'une population malade. Si un test diagnostic est positif, est-on malade pour autant? Le théorème de Bayes permet de calculer la probabilité d'avoir une maladie si le test est positif : c'est la **valeur prédictive positive** (VPP). Grâce à une méthode graphique, la VPP peut s'exprimer en fonction de la sensibilité  $Se$  du test, de sa spécificité  $Sp$ , et de la prévalence  $Pr$  de la maladie.



a) Un carré de côté 1 représente la population. La proportion  $Pr$  malade est colorée en rouge. b) Chez les malades, seule une proportion  $Se$  a un résultat de test positif (en rouge hachuré). Chez les bien portants, seule une proportion  $Sp$  a un test négatif (bleu non hachuré).

$$\text{VPP} = \frac{\text{Proportion de tests positifs et malades}}{\text{Proportion de tests positifs}} = \frac{\text{Aire rectangle } \blacksquare}{\text{Aire rectangle } \blacksquare + \text{Aire rectangle } \blacksquare}$$

$$\text{VPP} = \frac{Pr \times Se}{Pr \times Se + (1 - Sp) \times (1 - Pr)}$$

**DOC 6 Prévalence et valeur prédictive positive.****EXPLOITER LES DOCUMENTS**

- En quoi la détection de spam choisie est-elle une méthode bayésienne (**DOCS 1 et 2**) ?
- Quel sont les risques et les avantages à fixer une probabilité seuil très élevée ou au contraire très basse (**DOC. 2**) ?
- Calculer la fréquence de faux positifs et de faux négatifs, ainsi que la sensibilité et la spécificité dans la reconnaissance de spams (**DOCS 3 à 5**).
- Pourquoi le calcul de la valeur prédictive positive (VPP) fait-il appel au théorème de Bayes (**DOC. 6**) ?
- Comment définiriez-vous la valeur prédictive négative (VPN), probabilité d'être bien portant si le test est négatif (**DOC. 6**) ?
- Dans le cadre du dépistage du cancer de la prostate par dosage de PSA, calculez la VPP (**DOC. 7**).

Un test diagnostique rassemble des observations cliniques pour effectuer un diagnostic par exemple médical. On caractérise sa performance à l'aide de deux grandeurs.

- La sensibilité mesure la probabilité qu'un test soit positif lorsqu'il doit l'être, c'est-à-dire lorsque le critère à détecter (ex. : maladie, grossesse...) est bien présent. C'est un critère essentiel : on souhaite que le test détecte efficacement un état médical.
- La spécificité mesure la probabilité qu'un test soit bien négatif lorsqu'il doit l'être, c'est-à-dire lorsque le critère à détecter est absent. On souhaite que le test ne détecte rien s'il n'y a rien.

**ESPRIT CRITIQUE**

La prostate est une glande de l'appareil reproducteur masculin. Le cancer de la prostate touche une proportion notable des hommes, surtout après l'âge de 70 ans. On peut estimer la prévalence dans la population masculine de plus de 70 ans à environ 9 %. On le dépiste habituellement par le dosage dans le sang d'une protéine spécifique de la prostate, le PSA : une concentration supérieure à un seuil fixé peut être interprétée comme un indice d'une éventuelle tumeur. Si l'association française d'urologie « recommande une détection précoce individualisée du cancer de la prostate » (rapport de 2013), la Haute autorité de santé ne préconise pas un dépistage systématique par dosage du PSA (avis de février 2017).

Sensibilité	Spécificité
50 %	90 %

**DOC 7 Le dépistage du cancer de la prostate.****ESPRIT CRITIQUE**

En estimant à 4 millions les hommes de plus de 70 ans en France, pour combien d'hommes un diagnostic erroné sera-t-il posé ? Même question pour un diagnostic correct.

→ Commentez les avis divergents entre l'Association française d'urologie et la Haute autorité de santé. Un seul test médical vous semble-t-il suffisant pour établir le diagnostic d'un cancer ?

**Pistes de travail ► DOC. 7**

# L'INTELLIGENCE ARTIFICIELLE



## 1. Le traitement de l'information numérique

- ▶ Jusqu'au début du xx<sup>e</sup> siècle, les machines automatisées sont limitées à une ou quelques applications particulières. Turing est le premier à imaginer une **machine universelle** capable de traiter n'importe quel type de données. Cette machine est matérialisée 10 ans plus tard avec les premiers ordinateurs, composés au minimum d'un processeur et d'une mémoire vive. L'informatique se répand ensuite largement avec l'**informatique embarquée**. > **Unité 1**
- ▶ Les ordinateurs manipulent divers types de données numérisées : images, sons, vidéos... Les programmes sont des informations stockées dans le même type de mémoire que les données. Un programme écrit dans un langage de programmation de haut niveau, comme Python, peut être traduit en instructions spécifiques à chaque type de processeur. Ainsi, un **fichier exécutable** sur une machine n'est pas obligatoirement exécutable sur une autre. > **Unité 2**
- ▶ Un programme complexe moderne peut comporter plusieurs centaines de millions de lignes de code, ce qui rend très probable la présence de **bugs** (ou bogues). Ces erreurs peuvent conduire un programme à avoir un comportement inattendu et entraîner des conséquences graves. > **Unité 3**

## 2. Intelligence artificielle et apprentissage machine

- ▶ L'**intelligence artificielle** (IA) regroupe des techniques permettant à des machines de réaliser des actions normalement associées à l'intelligence humaine. L'**apprentissage machine** (ou automatique) est une branche de l'IA qui utilise une grande quantité de données (big data) pour entraîner un programme à repérer des corrélations ou des similitudes. Le programme, une fois entraîné correctement, est capable de prédire ou de décider à partir de données nouvelles. > **Unité 4**
- ▶ L'IA soulève de nombreuses questions éthiques. L'apprentissage machine utilise une grande quantité de données personnelles pour lesquelles un consentement explicite doit être obtenu. Par ailleurs, la qualité des données d'entraînement est essentielle puisque des biais en entrée peuvent se retrouver amplifiés en sortie. Le développement des voitures autonomes soulève d'autres questions éthiques, comme la responsabilité engagée lors d'un accident. > **Unités 5 et 6**

## 3. L'inférence bayésienne dans l'apprentissage automatique

- ▶ L'**inférence bayésienne** permet d'estimer la plausibilité d'une hypothèse, étant donné des observations concrètes. Elle est très utilisée en apprentissage automatique, notamment en vue de prononcer un diagnostic (médical, détection de spam...). Lorsqu'on applique une procédure de décision, nécessairement imparfaite, il apparaît des erreurs qu'on appelle des « faux », et qu'on qualifie de positif ou négatif selon le résultat de la détection. La construction d'un tableau de contingence permet de calculer la fréquence de faux positifs et de faux négatifs, qui caractérise la qualité de la procédure de décision.
- ▶ La performance d'un test diagnostique est de plus caractérisée à l'aide de la sensibilité, qui mesure la probabilité qu'un test soit positif lorsqu'il doit l'être, et de la spécificité, qui mesure la probabilité qu'un test soit négatif lorsqu'il doit l'être. Ces grandeurs permettent notamment de calculer la probabilité d'avoir une maladie si le test est positif : c'est la valeur prédictive positive dont le résultat dépend de la **prévalence** de la maladie. > **Unité 7**

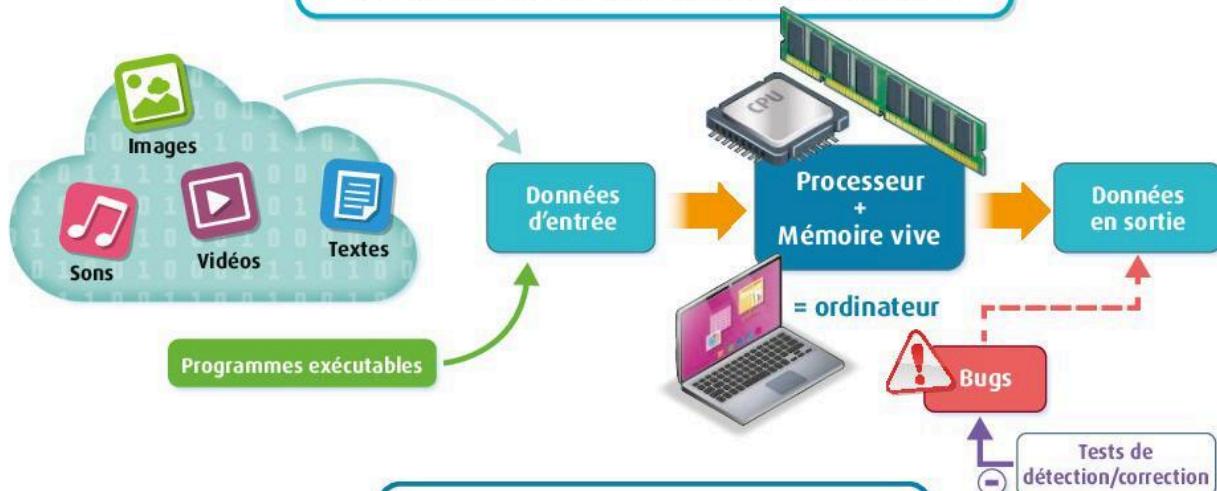
### Les mots-clés du chapitre

- **Machine universelle** : Machine, imaginée par Turing, capable de simuler n'importe quelle machine réalisant un calcul. La configuration nécessaire fait alors partie des données d'entrée.
- **Informatique embarquée** : Appareils de la vie courante dont le fonctionnement repose sur un ou plusieurs programmes informatiques.
- **Fichier exécutable** : Fichier qui contient le programme sous une forme qui peut être traduite en instructions spécifiques à chaque type de processeur.
- **Bug** : Erreur dans un programme entraînant un comportement inattendu lors de l'utilisation du programme.
- **Intelligence artificielle** : Ensemble de théories et de techniques qui traite de problèmes dont la résolution fait normalement appel à l'intelligence humaine.
- **Apprentissage machine** : Branche de l'IA qui utilise des programmes repérant des corrélations ou des similitudes sur un très grand nombre de données et peuvent, après cet entraînement, prédire des résultats ou prendre des décisions sur d'autres données.
- **Inférence bayésienne** : Méthode de calcul de probabilités de causes à partir des probabilités de leurs effets.
- **Prévalence** : Proportion de personnes malades au sein d'une population.

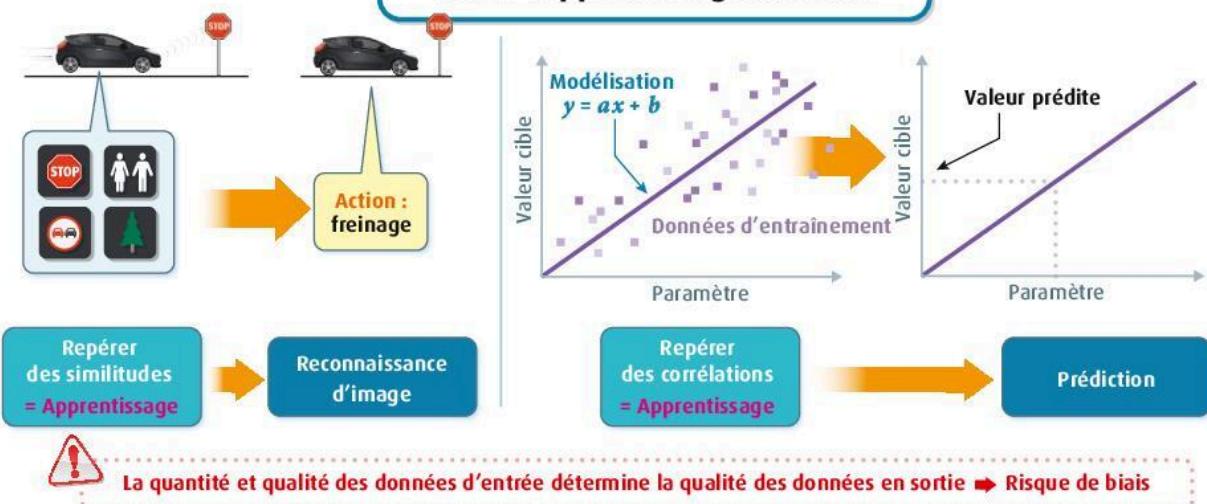
# *l'essentiel par l'image*

Schéma interactif 

## Le traitement de l'information numérique



## L'IA et l'apprentissage machine



## L'inférence bayésienne

► Méthode de calcul de la probabilité des causes à partir de la probabilité de leurs effets

► Tableau de contingence :

	Caractère présent	Caractère absent	
Test positif	Vrai positif (VP)	Faux positif (FP)	$f(FP) = \frac{FP}{VP + FP}$
Test négatif	Faux négatif (FN)	Vrai négatif (VN)	$f(FN) = \frac{FN}{FN + VN}$
	$\text{Sensibilité} = \frac{VP}{VP + FN}$	$\text{Spécificité} = \frac{VN}{FP + VN}$	

► Caractérisation de la qualité d'une procédure de décision dans un test diagnostique

## Mémoriser son cours

Exercices corrigés



Pour mémoriser l'essentiel du cours, posez-vous régulièrement ces questions et vérifiez vos réponses.

1. Citez trois noms de scientifiques connus pour leur rôle dans l'histoire de l'informatique et des ordinateurs.
2. Dans l'architecture de Von Neumann, quels sont les deux constituants essentiels d'un ordinateur?
3. Comment calcule-t-on la taille en octets d'un fichier texte?
4. Donnez l'ordre de grandeur de la taille de fichiers audio, image et vidéo.
5. Quelles peuvent être les conséquences d'un bug?
6. Comment une IA peut-elle prévoir une valeur à partir de paramètres donnés?
7. Pourquoi certains usages de l'IA posent-ils des problèmes éthiques?
8. Comment un test médical imparfait permet-il de détecter correctement une anomalie?

## Pour s'échauffer

Exercices interactifs corrigés



### 1 QCM

Pour chaque proposition, identifiez la (ou les) bonne(s) réponse(s).

**1. La place en mémoire d'une vidéo :**

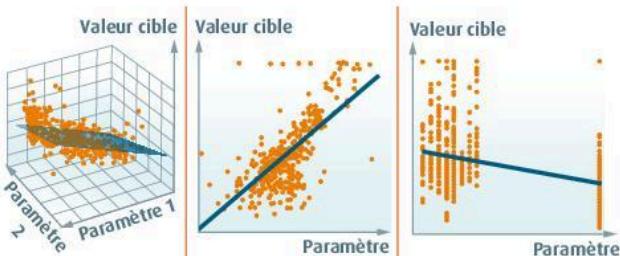
- a. est identique à une photo de même résolution.
- b. est supérieure à une page de texte.
- c. est inférieure à un son de même durée.
- d. est supérieure à une photo de même résolution.

**2. Pour entraîner une IA fondée sur l'apprentissage machine :**

- a. la quantité des données d'entraînement est importante mais pas la qualité de ces données.
- b. la qualité des données d'entraînement est importante mais pas la quantité de ces données.
- c. la qualité et la quantité des données d'entraînement est importante.
- d. ce qui importe c'est la qualité des données en sortie.

**3. Le modèle le plus en accord avec les données expérimentales est :**

- a.  $R^2 = 0,67$ .
- b.  $R^2 = 0,46$ .
- c.  $R^2 = 0,11$ .



**4. Un fichier comportant l'extension .exe**

- a. est un fichier exécutable sous windows.
- b. est un fichier exécutable sous MacOSX.
- c. est un fichier de données sous windows.
- d. est forcément un fichier à risque.

**5. On écrit en langage Python la fonction ci-dessous. Dans quel cas, cette fonction risque-t-elle de poser problème ?**

```
def arrondir_fraction(num,den) :
    return num / den
```

- a. arrondir\_fraction(0,5).
- b. arrondir\_fraction(5,0).
- c. arrondir\_fraction(1.5,5).
- d. arrondir\_fraction(5,2).

**6. Un test de reconnaissance du caractère manuscrit «3» donne le tableau de contingence suivant. Que peut-on dire de ce test ?**

	Le chiffre testé est un trois	Le chiffre testé n'est pas un trois
Test Positif	580	30
Test Négatif	35	5 355

- a. Il est parfait car il détecte tous les chiffres trois présentés.
- b. la fréquence de faux négatifs vaut  $\frac{35}{35 + 5 355} = 0,6\%$ .
- c. Il est imparfait mais ses résultats permettront de calculer le nombre de chiffres trois présents sur un échantillon futur en tenant compte de ses imperfections.
- d. 35 résultats sont des faux positifs.

**7. Lors de l'apprentissage machine :**

- a. les biais en entrée peuvent être amplifiés en sortie.
- b. des problèmes éthiques peuvent être soulevés.
- c. les données personnelles sont utilisées sans restriction.
- d. l'entraînement repose sur des corrélations.

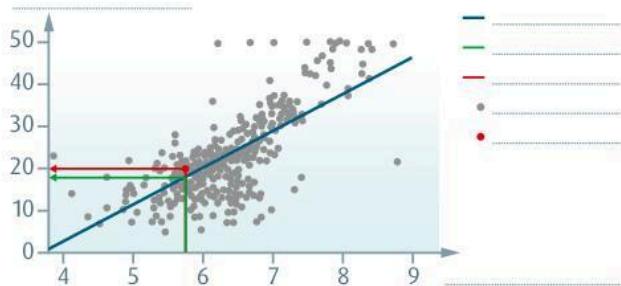
► CORRECTION p. 322

## 2 Vrai/Faux

Identifiez les affirmations fausses et rectifiez-les.

- Tous les systèmes informatiques, embarqués ou non, ont besoin d'être programmés par l'utilisateur.
- Un ordinateur doit comporter au moins un processeur et une mémoire vive.
- Lorsque l'on exécute un programme, du code écrit en langage machine est converti en code compréhensible par le programmeur (comme Python par exemple).
- Un ordinateur peut manipuler des données de natures diverses.
- Certains programmes peuvent comporter plusieurs millions de lignes de code.
- Il est très facile de supprimer tous les bugs d'un programme informatique.
- La spécificité mesure la probabilité qu'un test soit bien positif lorsqu'il doit l'être.
- Avec un test très sensible, on observe une fréquence faible de faux négatifs.

## 3 Apprentissage machine

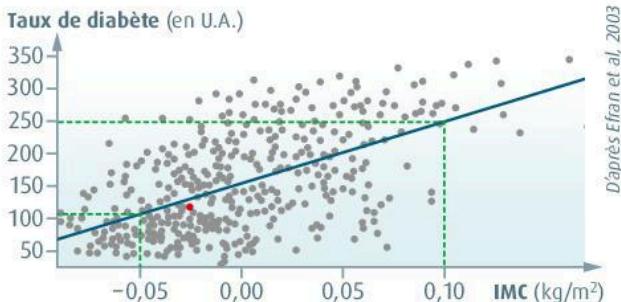


Lédez le graphique ci-dessus en utilisant les termes :

- points expérimentaux de l'ensemble d'entraînement
- courbe modèle
- valeur cible réelle
- valeur cible prévue par le modèle
- point expérimental de l'ensemble de test
- paramètre étudié
- valeur cible

## 4 Diabète et IMC

Grâce à une IA, on modélise l'influence de l'indice de masse corporelle (IMC) sur le taux de diabète (valeur cible). Cette modélisation a été réalisée à partir d'une base de données publique.



- Utilisez les lignes en pointillés du graphique ci-dessus afin de déterminer l'équation de la droite modèle (bleue) obtenue.
- Déterminez graphiquement puis en utilisant l'équation du modèle la valeur cible donnée par ce modèle dans le cas du point rouge de la figure.

## 5 Distance de freinage

À 50 km/h la distance d'arrêt d'un véhicule est d'environ 25 m.



**Question:** Argumentez sur l'apport des technologies de conduite autonome au niveau de la sécurité.

## 6 Test diagnostique

On note A la proposition «la personne est malade» et B la proposition «le test est positif».

- Écrivez en français ce que signifient  $p(A)$ ,  $p(B)$ ,  $p(A)$  connaissant B et  $p(B)$  connaissant A.
- À partir de cet exemple, écrivez en français ce qu'est un faux positif, un faux négatif, un vrai positif et un vrai négatif.
- Dans le(s)quel(s) des 4 cas ci-dessus la personne est-elle malade?

## 7 Algorithme

On considère l'algorithme suivant :

Soit L une liste de  $n$  éléments

$m = 1$

Pour  $i$  allant de 1 à  $n$

Si le  $i^{\text{ème}}$  élément de L est inférieur ou égal au  $m^{\text{ème}}$  élément de L

Alors stocker la valeur de  $i$  dans  $m$

Afficher le  $m^{\text{ème}}$  élément de L

- Que fait cet algorithme? Vous pouvez le tester avec la liste d'éléments [1, 3, 2] pour vous en rendre compte.
- Proposez deux listes d'éléments permettant de tester son comportement en expliquant ce que chacune permet de tester.
- Quel cas non prévu pourrait donner lieu à un bug ici?

## 8 Détection de spams

Après l'avoir entraîné à détecter des spams, on applique un algorithme à la détection de spams sur 5 216 courriels.

	Spam	Courriel normal
Test Positif	1200	84
Test Négatif	507	3 425

- Calculez la fréquence de faux positifs et de faux négatifs.
- Rappelez la définition de la spécificité et de la sensibilité d'un test.
- Calculez la spécificité et la sensibilité de la détection de spams.

## 9 Comparer des stockages de données numériques PIX

### Stocker des données texte sur son téléphone



Les smartphones de 2020 proposent des capteurs photo de 108 Mégapixels (millions de pixels) permettant de filmer en haute résolution (appelée 8K) à 30 images par seconde.

Une page de texte comporte en moyenne 2 300 signes, une photo à cette résolution prend environ 10 Mo pour être stockée et une minute de film en 8K occupe environ 1 000 Mo.

En prenant en photo la page d'un livre qui ne comporte que du texte, un élève se demande sous quelle forme les 200 pages de son livre prendraient le moins de place sur la carte mémoire de son téléphone.

#### QUESTIONS

- D'après vos connaissances, classez les trois formats proposés du plus économique en espace de stockage au plus gourmand.
- En considérant un codage ASCII, déterminez par le calcul l'espace occupé en mémoire sous format texte pour un livre de 200 pages. Donnez le résultat en Mbits et en Mo.
- Quel serait l'espace occupé en prenant tout le livre en photo à l'aide d'un smartphone (1 page par photo) ?
- Comparez les résultats précédents avec l'espace occupé par un film du livre à raison d'une page par seconde.

On considère que ces caractères sont codés en ASCII : un caractère est codé sous 8 bits = 1 octet.  
 $1 \text{ Go (gigaoctet)} = 10^3 \text{ Mo (mégaoctets)}$   
 $= 10^6 \text{ Ko (kilooctets)}$   
 $= 10^9 \text{ octets}$

#### RÉSOLUTION

- Le format texte est le plus économique en espace de stockage, vient ensuite la photo puis la vidéo.  
 On considère un octet occupé par caractère. Il y a 200 pages avec en moyenne 2 300 signes par page soit un total de  $200 \times 2300 = 460\,000$  signes.  
 L'espace occupé est donc de 460 000 octets soit 460 Ko ou encore 0,46 Mo.  
 Comme un octet vaut 8 bits, l'espace occupé est de  $0,46 \times 8 = 3,66$  Mbits.
- Avec un capteur de 108 Mégapixels, une photo prend 10 Mo. Il y a 200 pages donc 200 photos à prendre, le livre occuperait donc  $200 \times 10 = 2\,000$  Mo.
- Il y a 200 secondes de film à réaliser. Une minute de film occupe 1 000 Mo donc 1 seconde de film occupe environ  $\frac{1000}{60} = 17$  Mo. Pour 200 pages, on arrive à  $200 \times \frac{1000}{60} = 3\,333$  Mo.

Les résultats obtenus par ces calculs sont bien en accord avec mes connaissances.

Il faut bien utiliser une même unité pour pouvoir comparer les trois quantités de données.

Il est préférable de réutiliser 1 000 final plutôt qu'un résultat intermédiaire arrondi afin de limiter les erreurs.

Ne pas oublier de confronter les résultats numériques avec vos connaissances afin de détecter d'éventuelles erreurs de calcul. Si vous détectez une erreur et que vous ne trouvez pas d'où elle peut venir, il sera toujours plus profitable de l'indiquer : cela montre que la moitié du chemin vers la correction a déjà été faite.

## Exercices d'application Méthode

### 10 Comparer des stockages de données numériques PIX

#### Stockage sur cartes perforées

Les cartes perforées sont des dispositifs de stockage de texte (utilisés pour le stockage de programmes informatiques) brevetés en 1928 par IBM. Elles sont constituées de papier rigide permettant de coder sous forme binaire (trou ou absence de trou) chaque caractère sur une des 80 colonnes de la carte. Une carte permet le codage d'une ligne de programme. Chaque carte pèse 2,55 g, a une surface de 0,0155 m<sup>2</sup> et permet au maximum de stocker 960 bits (12 bits × 80 colonnes).



**DOC1** Cartes perforées d'IBM pour le stockage de texte.

#### AIDE

- 12 bits permettent  $2^{12}$  combinaisons différentes de 0 et de 1.



IBM travaille sur de nouveaux systèmes de stockage dont Millipede (mille-pattes en français) fondé sur le principe des cartes perforées mais miniaturisées. Le nom Millipede vient du fait que la machine qui réalise ces cartes de stockage utilise des milliers de « pointes » en silicium qui induisent chacune un bit de donnée sur une couche du matériau polymère. Un prototype présenté en 2005 permettait de stocker  $1,24 \cdot 10^{15}$  bits sur chaque mètre carré.

**DOC 2** Principe du système de stockage Millipede d'IBM.

#### QUESTIONS

- En observant la carte du **DOC. 1**, sur combien de bits sont codés les caractères?
  - Combien de caractères différents peuvent être codés?
  - Au 1<sup>er</sup> janvier 2020, le noyau Linux contenait 27,8 millions de lignes de code. Combien de cartes perforées auraient été nécessaires pour stocker ce programme?
  - En déduire la masse et la surface de papier nécessaires pour stocker ce programme sur ce support.
- Sachant que la Corse a une surface de 8 722 km<sup>2</sup>, commentez votre résultat.
- Quelle surface de cartes perforées faut-il pour stocker la même quantité d'informations que dans un mètre carré de cartes Millipede (**DOCS 1 et 2**)?
  - Combien de nanomètres carrés y a-t-il dans un mètre carré?
  - En considérant que chaque trou code 1 bit, déterminez l'ordre de grandeur de la taille d'un trou.

### 11 Comparer des stockages de données numériques PIX

#### Différents supports de stockage

Support de stockage	Capacité	Support de stockage	Capacité
Disque dur	1 To	CD ROM	700 Mo
Clé USB	32 Go	Disquette*	1,44 Mo

\*support de stockage inventé en 1967 par IBM, la disquette est progressivement remplacée par les clés USB et les disques durs dès la fin des années 90.

#### QUESTIONS

- Donnez la capacité de chaque support en kilooctets.
- De combien de disquettes le disque dur et la clef USB sont-ils l'équivalent?
- De combien de CD ROM le disque dur et la clef USB sont-ils l'équivalent?
- On considère que la taille moyenne d'une photo est de 5 Mo et qu'une minute de film occupe environ 150 Mo.

Combien de photos, de minutes de films et de caractères (codés en ASCII) peuvent contenir ces 4 supports? Vous présenterez vos résultats sous forme de tableau en donnant le détail des calculs pour la première ligne de votre tableau.

## Tester ses compétences

### 12 Corriger un programme bogué PIX

#### Fonction moyenne en Python

On donne le programme Python ci-dessous.

##### QUESTIONS

1. Commentez les lignes 7 à 11 du programme.
2. Lorsque l'on exécute `moyenne([ ])` l'interpréteur nous renvoie `None`. Quelles sont les lignes de la fonction moyenne effectivement exécutées lors de cet appel?
3. Proposez une liste à passer à la fonction moyenne permettant de tester les autres lignes du programme ainsi que la sortie associée attendue.
4. Lorsque l'on exécute `moyenne('bonjour')` un message d'erreur est affiché. Comment s'appelle ce phénomène? Expliquez en français ce qu'il faudrait vérifier pour éviter le problème. Entre quelles lignes du programme doit être positionnée cette vérification?

```

1 def moyenne(liste) :
2     n = len(liste) # la variable n est affectée du nombre
3                     # d'éléments de liste
4     if n == 0 : # si n est égal à zéro
5         return None # Renvoie None puis met fin à l'exécution
6                     # de la fonction
7     else : # sinon
8         somme = 0 # on initialise la variable somme à zéro
9         for i in range(0, n): # Pour i entier de 0 à n-1
10             somme = somme + liste[i]
11     return somme / n

```

5. Une situation problématique apparaît lorsque l'on exécute `moyenne([True, False])` car l'interpréteur n'affiche pas un message d'erreur ici mais renvoie `0.5`. L'erreur est donc plus difficile à détecter. Comment s'appelle ce phénomène? Expliquer en français ce qu'il faudrait vérifier pour éviter le problème. Entre quelles lignes du programme doit être positionnée cette vérification?

#### AIDE

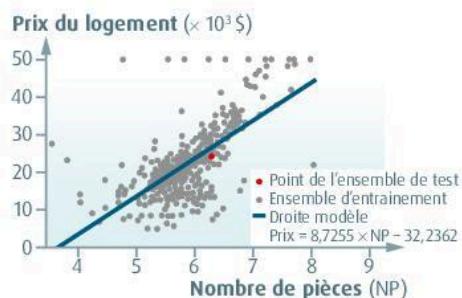
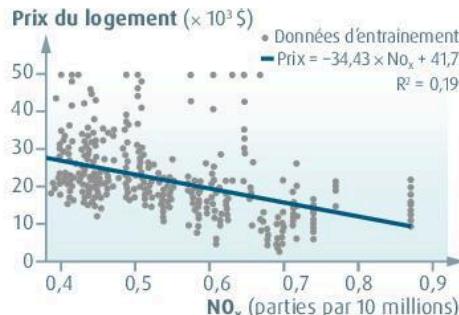
- La fonction moyenne attend une liste de valeurs. Elle exécute les lignes du bloc «if» si cette liste est vide et celles du bloc «else» sinon.

### 13 Analyser un exemple d'utilisation de l'IA PIX

#### L'IA pour calculer le prix d'un logement

On utilise une IA pour modéliser mathématiquement des données sur les prix des logements à Boston en utilisant différentes corrélations.

**DOC 1** Corrélation entre le prix du logement à Boston (valeur cible) et la concentration en polluants NO et NO<sub>2</sub> (appelés NO<sub>x</sub>).



**DOC 2** Corrélation entre le prix du logement à Boston (valeur cible) et le nombre de pièces moyen par appartement (NM).

On considère un point de l'ensemble de test (point rouge) qui a pour caractéristiques Prix = 23,7; NM = 6,75 et ÂGE = 74,9.

##### QUESTIONS

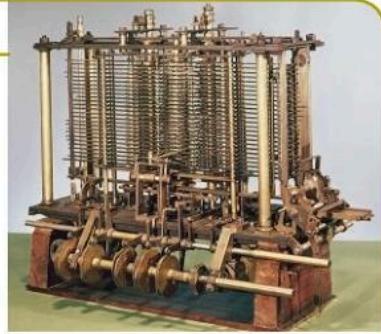
1. Pour chacun des documents, indiquez le signe de la pente de la droite de modélisation. Expliquez pourquoi ce résultat était prévisible.
2. La modélisation du doc.1 est-elle bien représentative des données? Répondez en utilisant deux arguments.
3. Quelle est la valeur du logement prédite par le modèle pour une concentration en NO<sub>x</sub> de 1 partie par 10 millions (DOC. 1)?
4. Déterminez par lecture graphique la valeur cible donnée par le modèle dans le cas du point rouge considéré (DOC. 2).

5. Déterminez alors par le calcul la valeur cible donnée par le modèle pour ce point. Quelle méthode vous semble être la plus précise (DOC. 2)?
6. On souhaite maintenant utiliser un modèle prenant en compte à la fois le nombre moyen de pièces NM et l'âge du logement ÂGE. Le modèle obtenu nous donne Prix = 8,39 × NM - 0,0702 × ÂGE - 25,3. Déterminez la valeur cible donnée par ce modèle pour le point rouge de l'ensemble de test (DOC. 2).
7. Lequel des deux modèles donne le meilleur résultat pour ce point (DOC. 2)?

## 14 Comprendre un algorithme de calcul PIX

### La machine de Babbage

La machine conçue par Charles Babbage en 1834 est destinée à automatiser les calculs de polynômes. Le but est d'obtenir des tables de valeurs pour certaines fonctions afin d'enlever les erreurs dues aux calculs manuels. La précision de la machine n'est cependant pas infinie, comme dans le cas des ordinateurs et calculatrices.



L'algorithme de fonctionnement de la machine de Babbage permet, une fois la machine réglée, de calculer un polynôme en ne réalisant que des additions. Voici un exemple pour le calcul de  $g(x) = x^3$ . On peut montrer (on l'admettra ici) que la valeur dans la colonne E est égale à 6 dans ce cas pour toutes les lignes à partir de la ligne 5.

	A	B	C	D	E
1	$x$	$g(x)$			
2	0	0			
3	1	1	$C_3 = B_3 - B_2 = 1$		
4	2	8	$C_4 = B_4 - B_3 = 7$	$D_4 = C_4 - C_3 = 6$	
5	3	27	$C_5 = B_5 - B_4 = 19$	$D_5 = C_5 - C_4 = 12$	$E_5 = D_5 - D_4 = 6$

Pour calculer la ligne suivante, il suffit de partir de la colonne E dans laquelle on inscrit 6 puis on se sert de ce résultat pour calculer la valeur dans la colonne D par addition, puis la valeur dans la colonne C par addition et enfin la valeur de  $g(x)$  dans la colonne B par addition.

6	4	$B_6 = C_6 + B_5$ $27 + 37 = 64$	$C_6 = D_6 + C_5$ $19 + 18 = 37$ On a aussi: $C_6 = B_6 - B_5$	$D_6 = E_6 + D_5$ $12 + 6 = 18$ On a aussi: $D_6 = C_6 - C_5$	$E_6 = D_6 - D_5 = 6$
---	---	-------------------------------------	---	--	-----------------------

#### AIDE

- On a toujours 6 dans la colonne E, que doit-on calculer en D7?
- La valeur attendue en C7 est 61.

#### QUESTIONS

- Écrivez en français l'algorithme permettant de calculer  $g(x)$  pour  $x$  supérieur à 3 en utilisant cette méthode. Vous pourrez utiliser des variables nommées A, B, C et D (comme les colonnes) initialisées avec les valeurs de la ligne  $x = 3$ . Utilisez cet algorithme (à la main) afin de calculer  $g(5)$  et vérifiez le résultat obtenu.
- Expliquez ce qu'apporte l'automatisation dans le contexte historique de l'invention de la machine de Babbage.
- Pour aller plus loin, codez l'algorithme de la question 1 en utilisant un tableur ou le langage de programmation Python. Déterminez ainsi  $g(100)$  et vérifiez la validité du résultat obtenu.

## 15 Utiliser un tableau de contingence

### Détection de cancers du sein par l'IA

Une IA est entraînée à reconnaître les cas de cancer du sein à partir d'une base de données publique (Breast cancer wisconsin (diagnostic) dataset).

#### QUESTIONS

- Déterminez les fréquences de vrais positifs, faux positifs, vrais négatifs et faux négatifs pour ce test (DOC.1).
- Quelle est la fréquence de l'événement « la personne est malade sachant que le test est positif » (DOC.1)?
- Calculez la sensibilité et la spécificité du test (voir DOC.5 p. 297).

	Individu Malade	Individu sain
Test positif	31	2
Test négatif	2	79

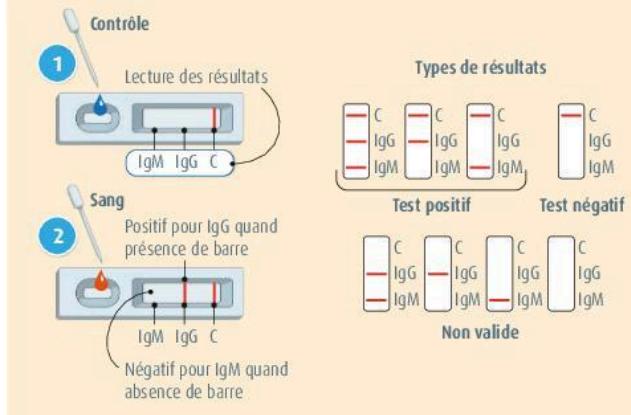
DOC1 Tableau de contingence du test de détection du cancer du sein.

## 16 Utiliser un tableau de contingence

## La détection de virus

Il existe différents tests permettant de détecter une infection virale chez les humains. Certains reposent sur la détection de l'ADN viral, d'autres sur la présence d'anticorps reconnaissant des protéines (antigènes) du virus.

On considère un test de détection virale consistant à observer dans le sang du patient la présence d'anticorps dirigés contre des protéines virales. Le but étant de détecter tous les malades, il a une sensibilité très élevée de 99,99 % et une spécificité de 94,1 %. Cette sensibilité est obtenue en combinant les résultats de deux détections : la détection d'anticorps appelés IgG et celle d'anticorps IgM dans le sang du patient. Le test sur les IgG seul a une sensibilité de 99,9 % et une spécificité de 98 %. La ligne C (« Contrôle ») du test détecte une protéine présente dans le sang humain et dans la solution contrôle.



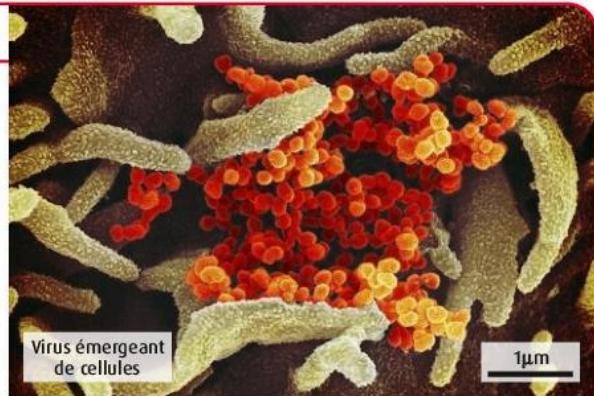
**DOC 1** Test de détection d'un virus à partir des anticorps sanguins.

## QUESTIONS

- Expliquez l'intérêt du contrôle.
- Sur l'échantillon de population à risque considéré, qu'on a diagnostiqué par une méthode de référence : quel est le nombre total de tests réalisés ? Le nombre P de tests positifs ? Le nombre N de tests négatifs ?

On note VP, VN, FP, FN le nombre de vrais positifs, vrais négatifs, faux positifs et faux négatifs respectivement. On note respectivement M et S le nombre de personnes malades et saines dans l'échantillon de population considéré. On note T le nombre total de patients de l'échantillon.

- Construisez un tableau de contingence puis exprimez N, P, M et S en fonction de VP, VN, FP et FN.
- Exprimez VP en fonction de la sensibilité Se et de M. Même question pour VN en fonction de la spécificité Sp et de S. En déduire que :
  - $FP = S \times (1 - Sp)$
  - $P = Se \times M + S \times (1 - Sp)$
  - $FN = M \times (1 - Se)$
  - $N = Sp \times S + M \times (1 - Se)$
- On obtient ainsi  $P = M \times Se + (T - M) \times (1 - Sp)$  d'où  $M = \frac{P - T \times (1 - Sp)}{Se + Sp - 1}$ . Calculez alors les valeurs de M et S. En déduire FN.
- Si on ne considère que le test sur l'IgG, que deviennent les valeurs de P, N, Se, et Sp ? Que devient la valeur de FN ? En déduire l'intérêt de faire un test couplé IgG, IgM.



68071	15341	6838	163083
C	C	C	C
IgG			
IgM			

**DOC 2 Résultats du test.** En réalisant ce test sur un échantillon de population à risque, on obtient les effectifs ci-dessus.

On rappelle que la sensibilité (Se) d'un test diagnostique mesure la probabilité qu'un test soit positif lorsqu'il doit l'être, c'est-à-dire lorsque la maladie est bien présente. La spécificité (Sp) d'un test diagnostique mesure la probabilité qu'un test soit négatif lorsqu'il doit l'être, c'est-à-dire lorsque la maladie est bien absente.

**DOC 3 Sensibilité (Se) et spécificité (Sp).**

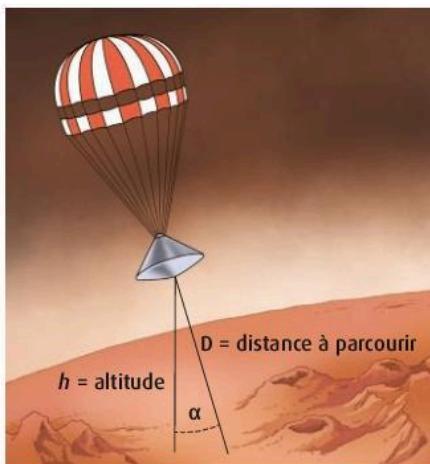
## 17 Corriger un programme bogué PIX

### Crash de l'engin spatial Schiaparelli

L'atterrisseur Schiaparelli conçu par l'Agence spatiale européenne s'est crashé sur le sol martien le 19 octobre 2016. Lors du déploiement du parachute, des oscillations imprévues de l'atterrisseur ont causé une saturation de l'unité de mesure d'inertie. Cette donnée non prévue au départ a causé une erreur lors du calcul de l'angle de l'atterrisseur d'environ 165 degrés – c'est à dire avec le bouclier thermique vers le haut – ce qui est impossible lors de l'utilisation du parachute. Cette donnée a ensuite été utilisée pour calculer l'altitude de l'atterrisseur, calcul donnant une valeur négative ! Ainsi les rétro fusées destinées à ralentir l'atterrisseur se sont arrêtées 34 secondes trop tôt causant un crash de la sonde à une vitesse de plus de 150 m/s.



#### DOC 1 Déroulement du crash.



```

1 def altitude(D, alpha):
2     if -90 < alpha < 90: # Si bouclier thermique vers le bas
3         angle_rad = alpha * pi / 180 # Conversion d'un angle
4             de degrés en radians
5         h = D * cos(angle_rad)
6         if h < 1. : # Si l'altitude est inférieure à 1 m
7             arret_moteurs( ) # Alors on coupe les moteurs
8             return h # On renvoie la valeur de h calculée
9     else: # Sinon
10        return None # On renvoie None car la valeur de h n'est
11            pas bonne

```

#### DOC 2 Correction du bug.

#### QUESTIONS

- Expliquez ce qu'est un bug.
- Quel est le bug rencontré ici ? Aurait-il été évitable (DOC. 1) ?
- Comment peut-on s'en protéger pour la future mission ExoMars 2020 (DOC. 1) ?
- Si on ajoute un commentaire dans le programme du DOC. 2, cela change-t-il son comportement ? Quel est donc l'intérêt de mettre un commentaire ?
- Si on ajoute un commentaire dans ce programme, cela change-t-il la quantité de mémoire occupée par le fichier ?
- Indiquez l'espace mémoire occupé par le programme sachant qu'il comporte 450 caractères.
- Ajoutez un commentaire à la ligne 4.

- altitude(10000, 45) renvoie environ 7071. Indiquez les lignes du programme effectivement exécutées lors de cet appel. Ce comportement est-il correct pour la fonction altitude ?
- Proposez des données d'entrée permettant de tester les lignes 8 et 9.
- Dans tous les cas précédents, la ligne 6 n'a pas été exécutée. Proposez des données d'entrée permettant de tester cette ligne.
- À votre avis, quelles lignes ont été rajoutées afin d'éviter le bug précédent ?
- Proposez une donnée d'entrée qui pourrait encore causer un bug ici.