

# Théorie de l'information

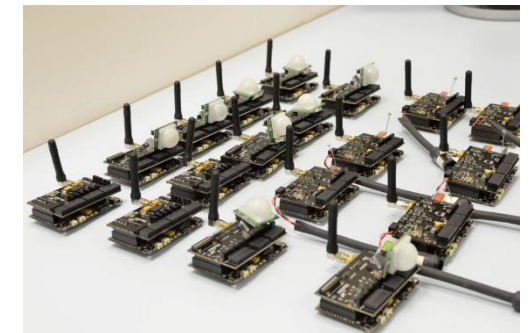
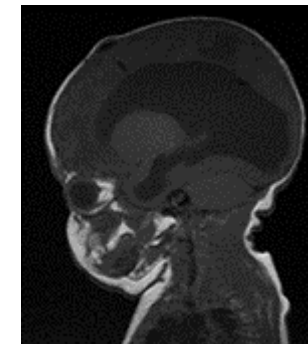
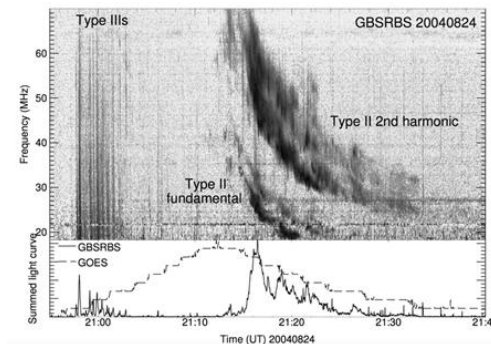
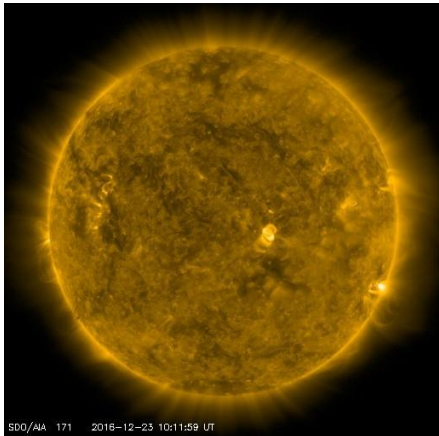


# Quelques présentations...

Nom : Adeline Paiement

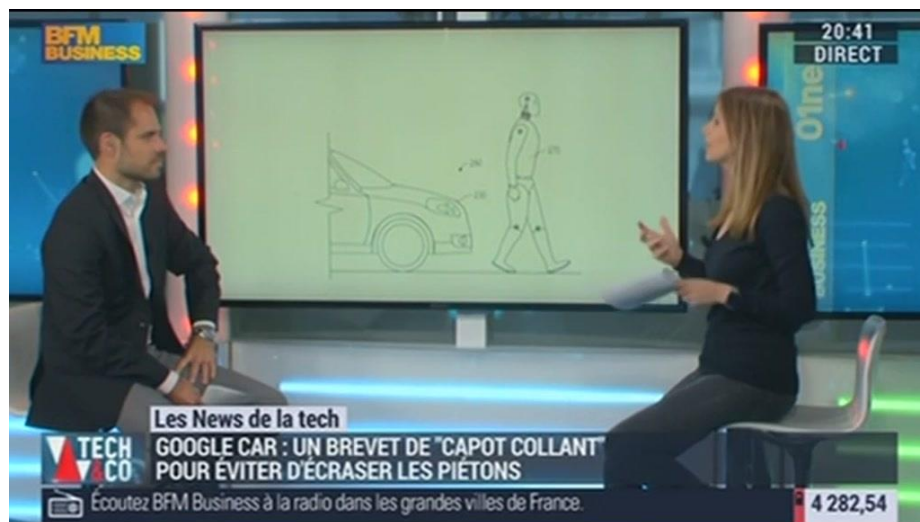
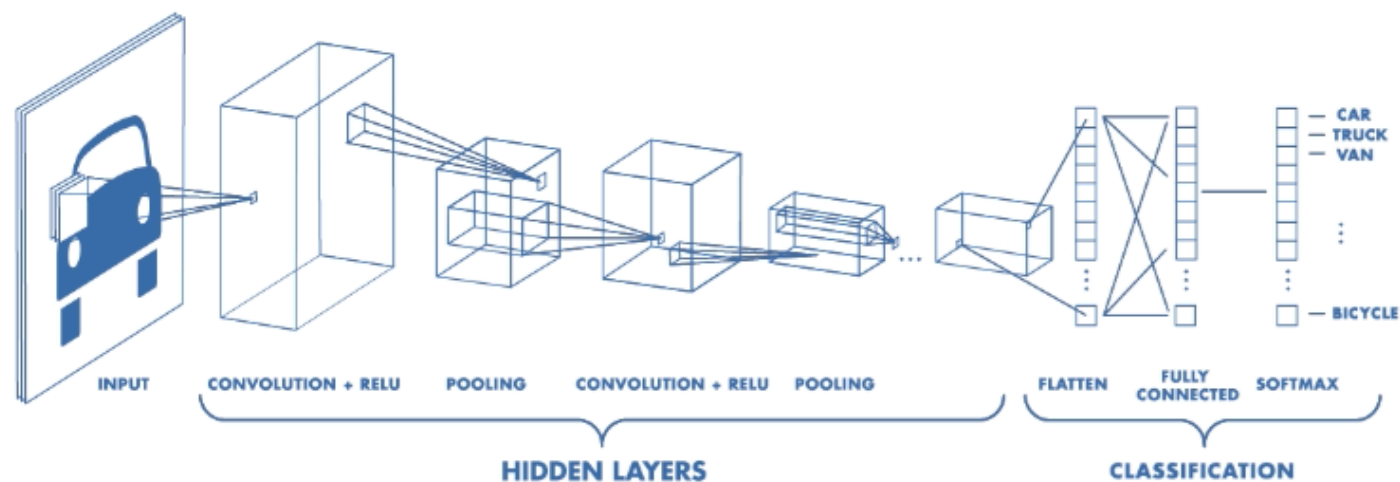
Métier : Enseignante-chercheure

Domaine de recherche : **IA, science des données**



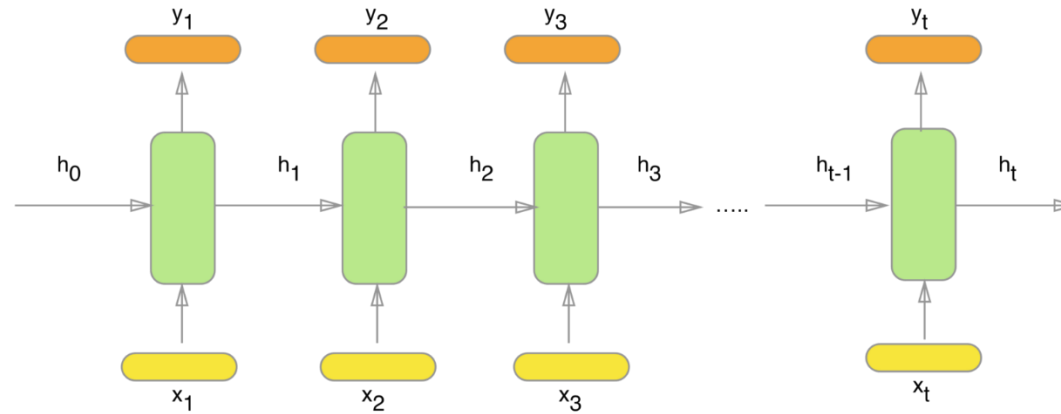
Qui dit IA, dit...

# Qui dit IA, dit...



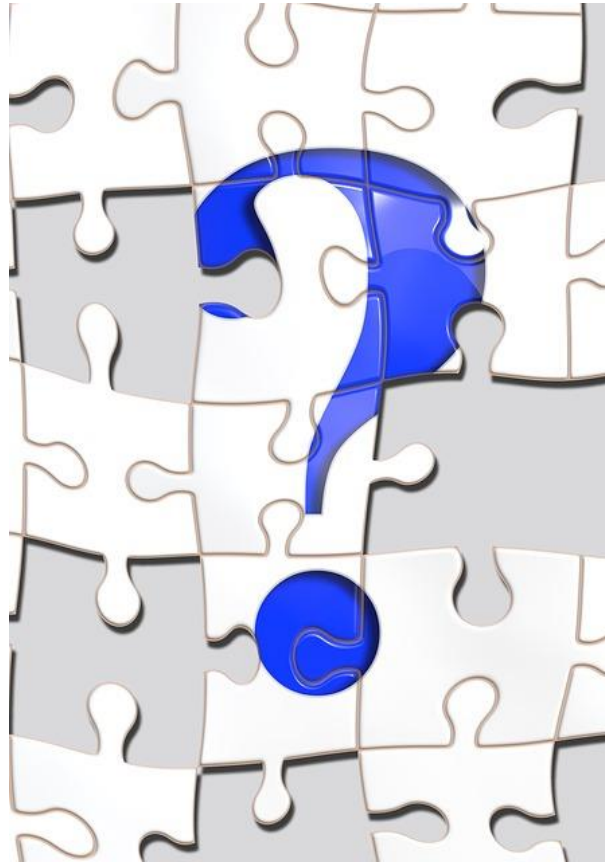
Qui dit science des données, dit...

# Qui dit science des données, dit...



Peut on vraiment résumer  
l'IA & la science des données  
à des réseaux de neurones ?

IA & science des données : qu'est ce que c'est (vraiment) ?

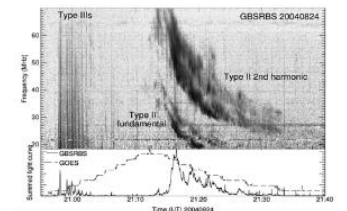
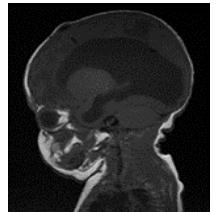
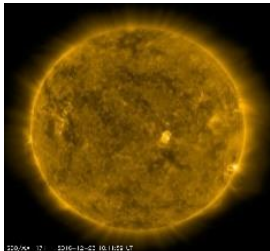
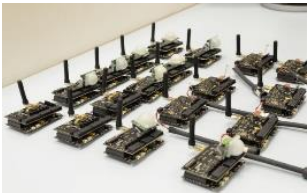


- Indice : cherchez le lien avec l'**information**...



# Qu'est ce que la science des données ?

Réponse : Extraire **l'information utile** cachée / noyée dans les données

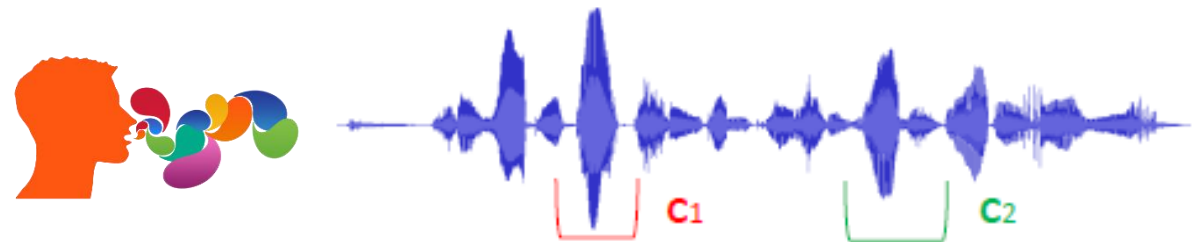


# Dans ce module :

Théorie de l'information...

...du point de vue de la science des données

...avec des applications concrètes

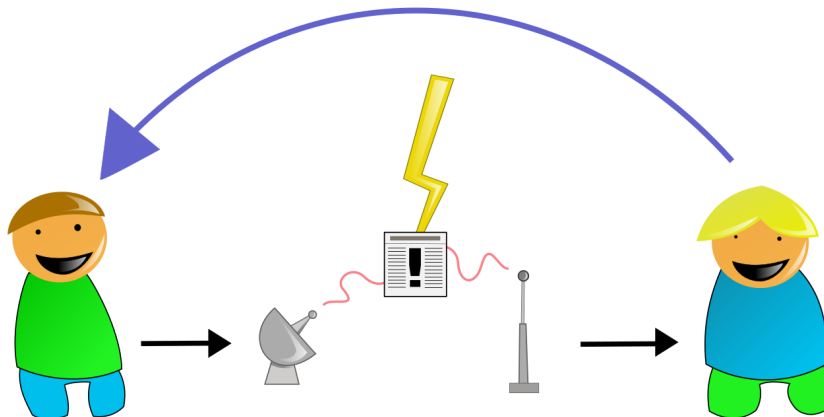


# Petit historique de la théorie de l'information

Boltzmann et Markov, fin XIX<sup>ème</sup> : probabilité d'un événement

Claude Shannon, 1949: *A Mathematical Theory of Communication*

- La Compagnie des Téléphones Bell cherche à envoyer des messages de manière plus économique et plus fiable
- Information comme grandeur mesurable
- Théorie *probabiliste* de la *transmission* de l'information



# Questions à aborder

## Théorie de l'information en science des données

- Qu'est ce que l'information en science des données ?
- Quelles propriétés et défis pour cette information ?

## Quelques rappels et définitions fondamentales

- Probabilités
- Quantité d'information
- Entropie
- Indépendance
- Information mutuelle

## Quelles applications en science des données ?

# Questions à aborder

## Manipulation des données

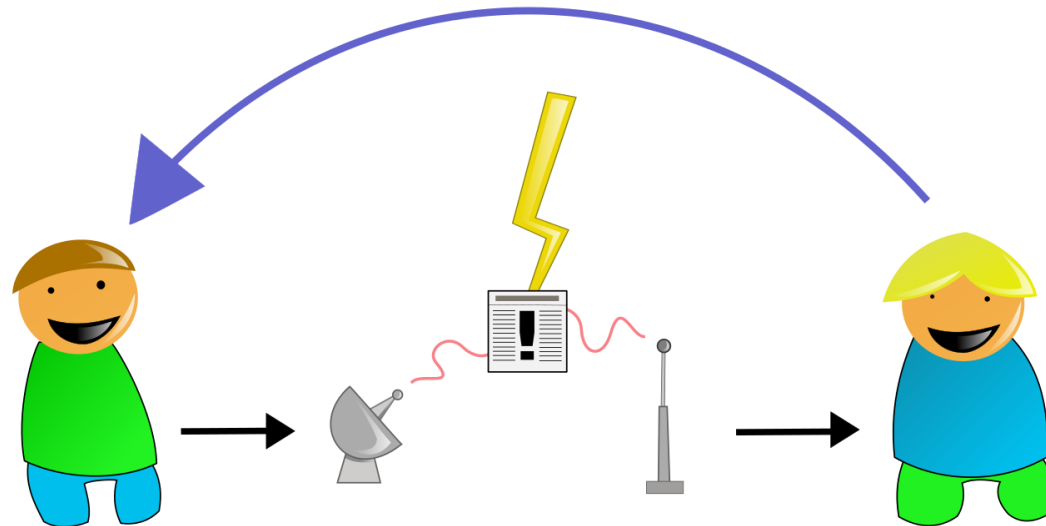
- Peut on transformer les données tout en préservant l'information ?
  - Nettoyage
  - Simplification
  - etc.

## Transmission de l'information

- Comment s'assurer qu'un message envoyé – reçu transporte bien l'information ?
- Application du principe à l'analyse des données (science des données)

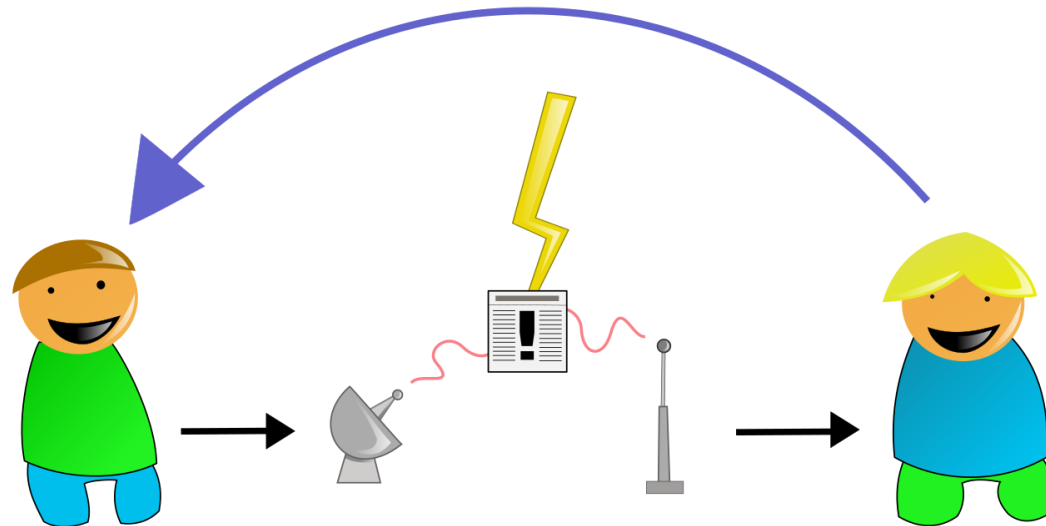
# Théorie de (la communication de) l'information

- Comment transmettre puis récupérer l'information utile via un message (les données transmises) pouvant être dégradé pendant la transmission ?



# Théorie de l'information *en science des données*

- Comment récupérer l'information utile  
des données mesurées (c.à.d. transmises)  
pouvant être dégradée pendant la mesure ?



# Qu'est ce que l'information ?

- Peut prendre beaucoup de formes



# Qu'est ce que l'information ?

- Peut prendre beaucoup de formes
- N'est pas observable directement, mais peut être *mesurée*

# Qu'est ce que l'information ?

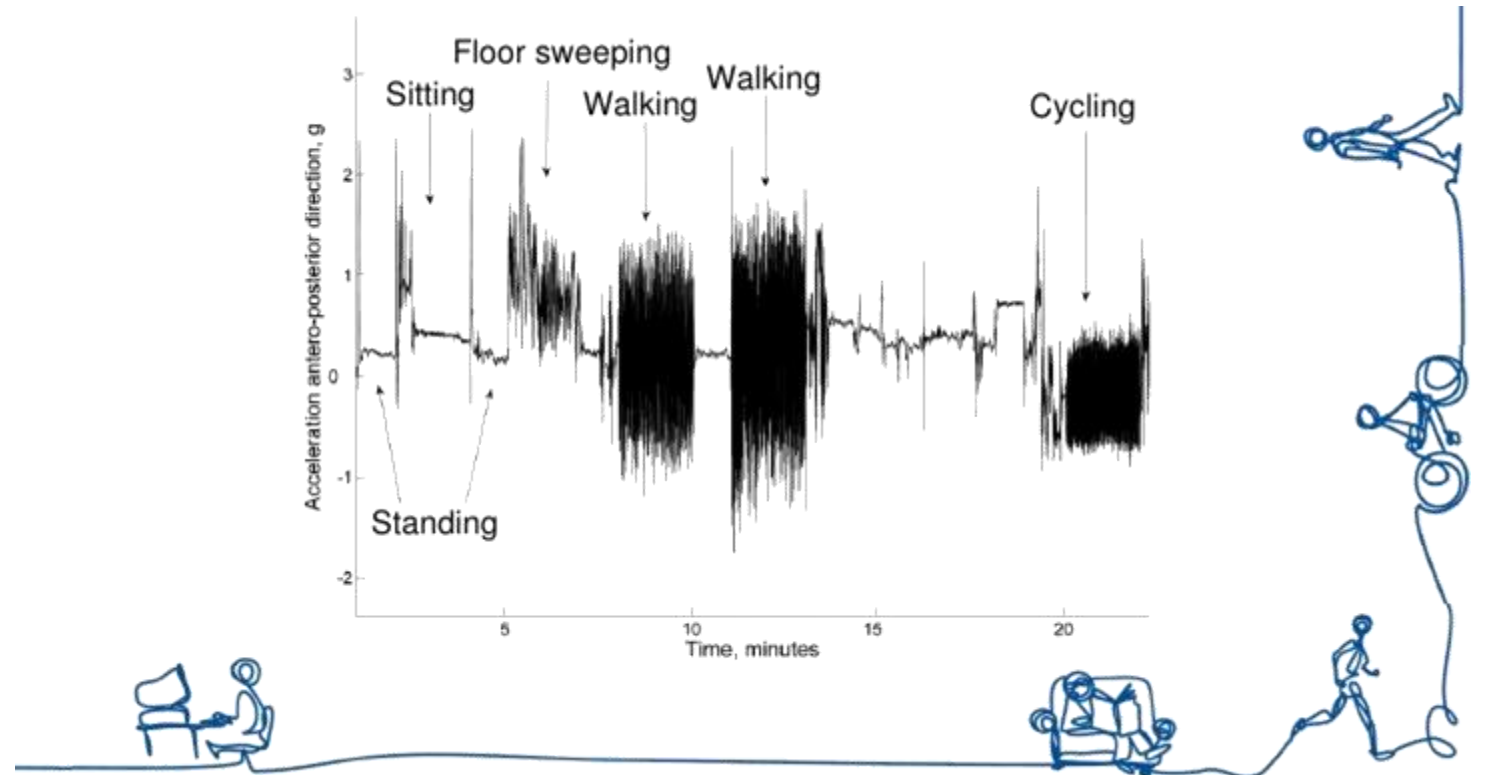
- Peut prendre beaucoup de formes
- N'est pas observable directement, mais peut être *mesurée*
- Peut être contenue dans des « messages » / données de types variés

# Qu'est ce que l'information ?

- Peut prendre beaucoup de formes
- N'est pas observable directement, mais peut être *mesurée*
- Peut être contenue dans des « messages » / données de types variés
- C'est le contenu **utile** du message / des données

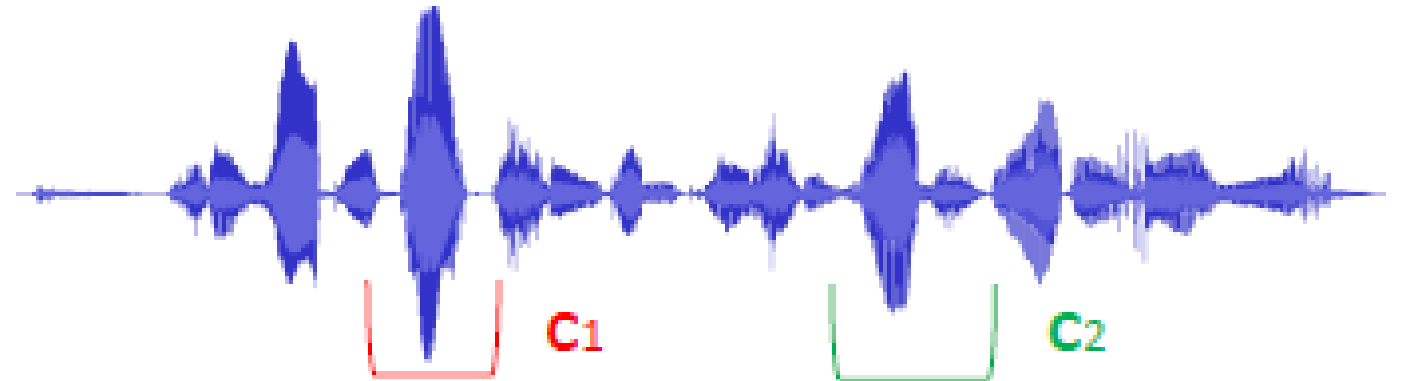
# Exemples de données et d'information

- Accélérométrie et reconnaissance d'activités



# Exemples de données et d'information

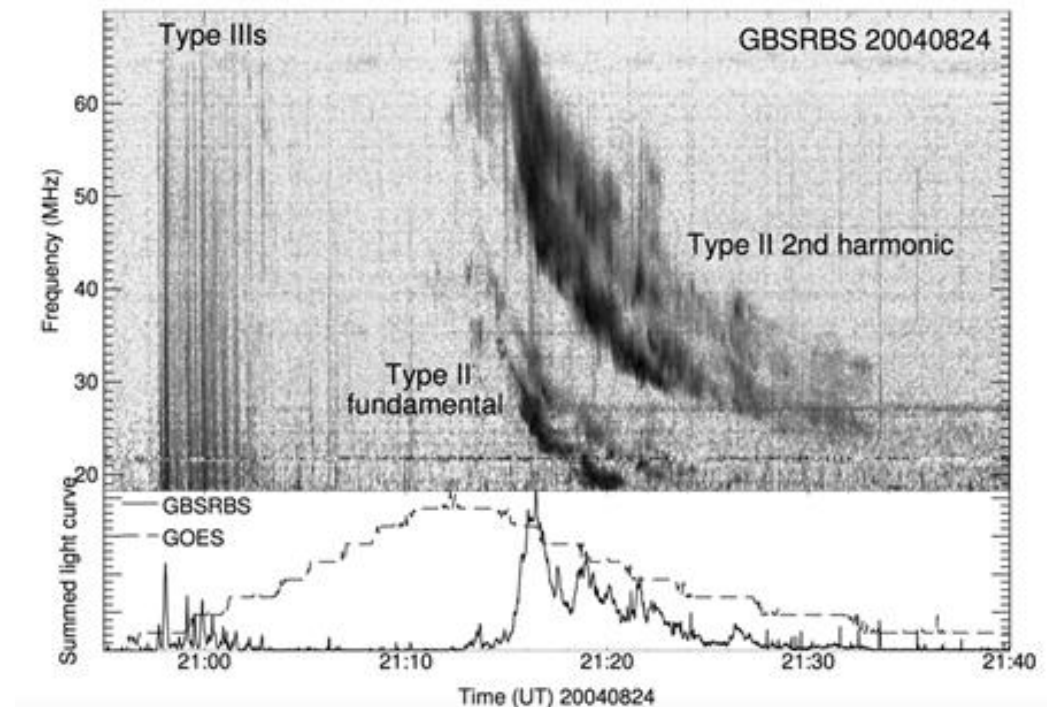
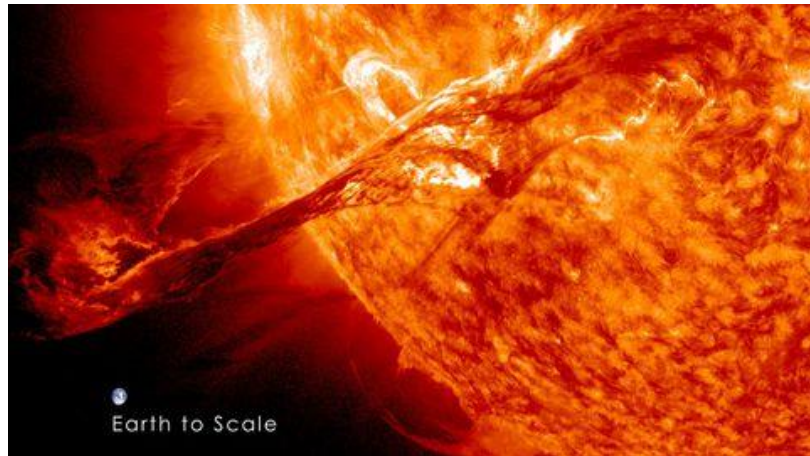
- Son et reconnaissance de la parole



Mot  
Signification de la phrase

# Exemples de données et d'information

- Emissions en fréquences radio et détection / caractérisation d'éruptions solaires

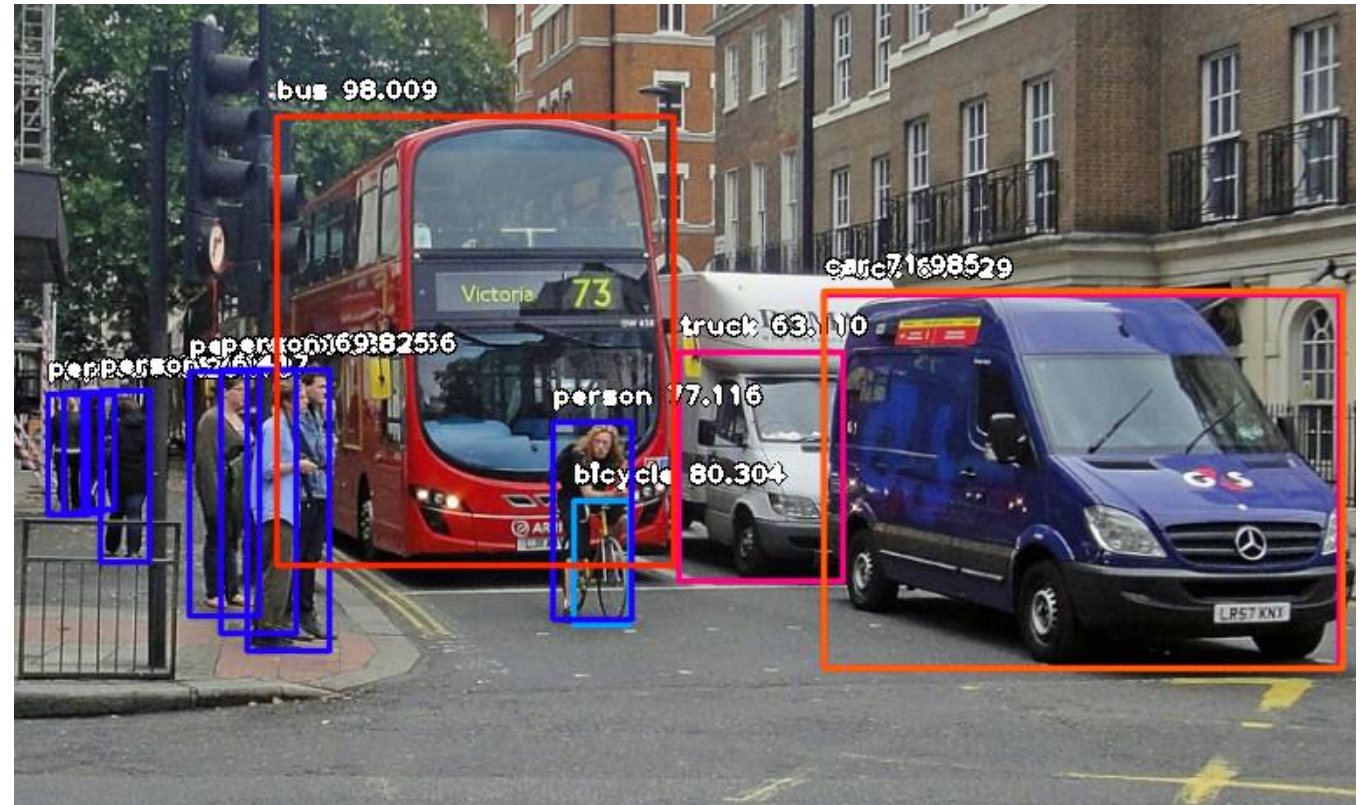


Position  
Durée  
Epaisseur  
Vitesse de décroissance



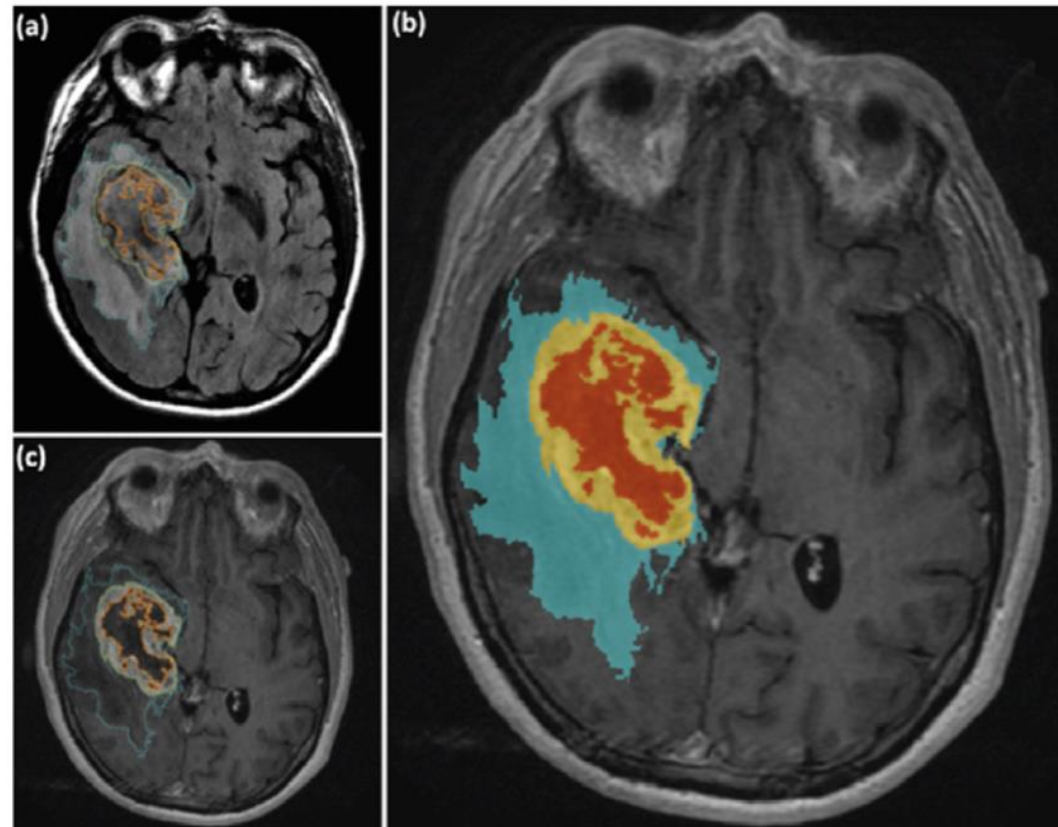
# Exemples de données et d'information

- Vidéo et détection / suivi de voitures et piétons



# Exemples de données et d'information

- IRM et détection / caractérisation de tumeurs



Position  
Etendue / taille  
Evolution au cours du temps

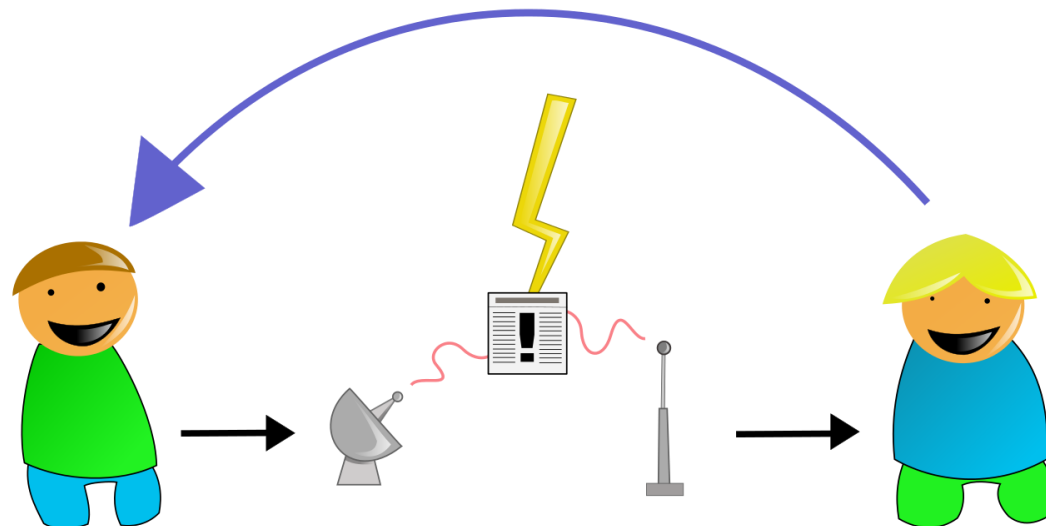


# Qu'est ce que l'information ?

- Peut prendre beaucoup de formes
- N'est pas observable directement, mais peut être *mesurée*
- Peut être contenue dans des « messages » / données de types variés
- C'est le contenu **utile** du message / des données
- Souvent noyée dans les données (autres « info » inutiles, redondances), et partiellement masquée par du bruit

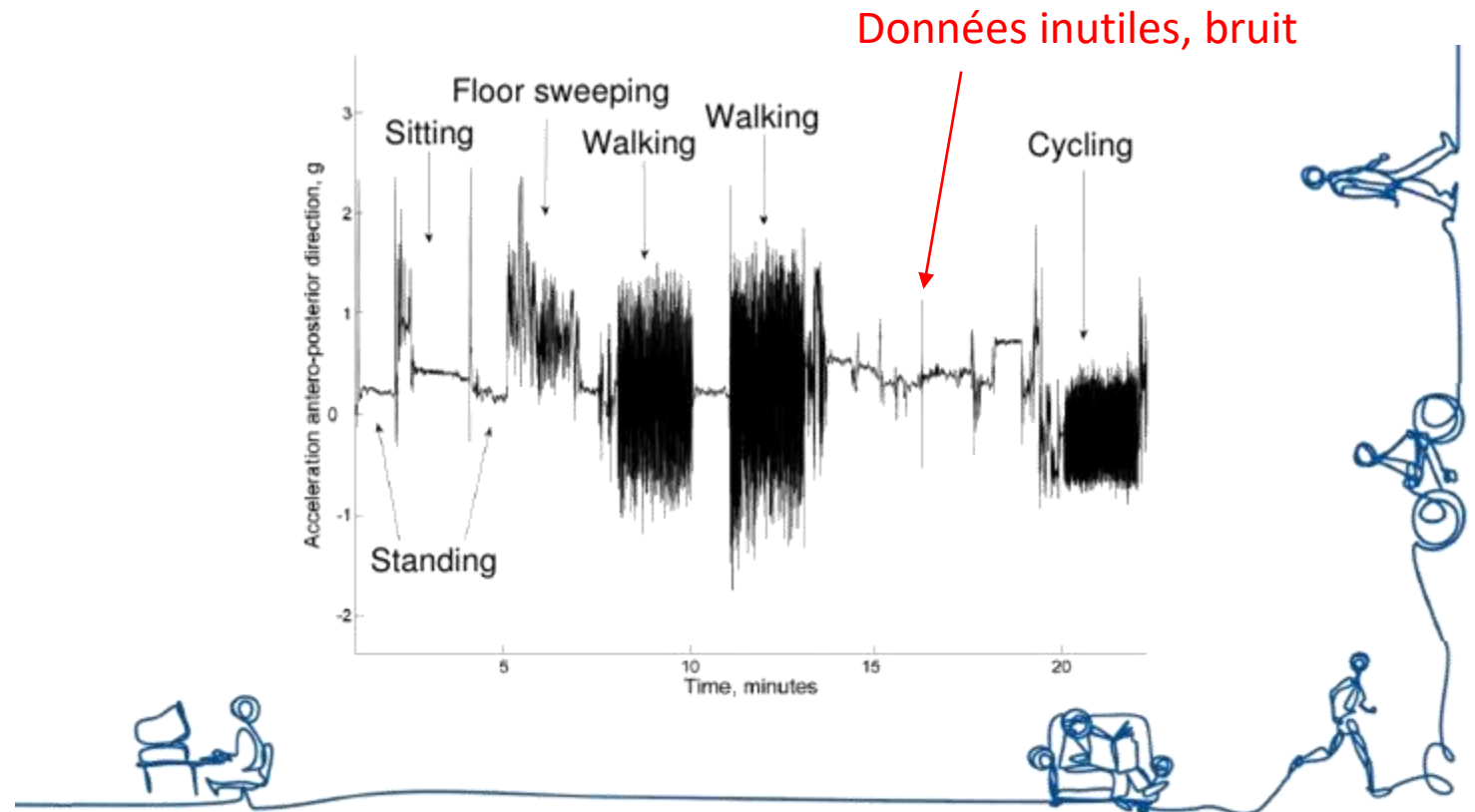
# Théorie de l'information en science des données

- Comment **recupérer** l'**information utile**  
des **données mesurées** (c.à.d. **transmises**)  
**pouvant être dégradée pendant la mesure ?**



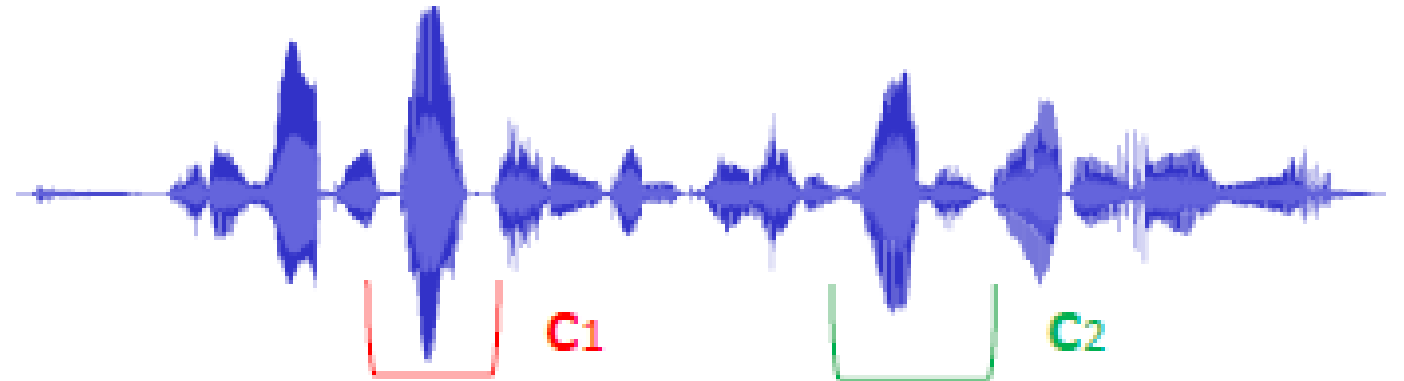
# Exemple de difficultés pour trouver l'information

- Accélérométrie et reconnaissance d'activités



# Exemple de difficultés pour trouver l'information

- Son et reconnaissance de la parole



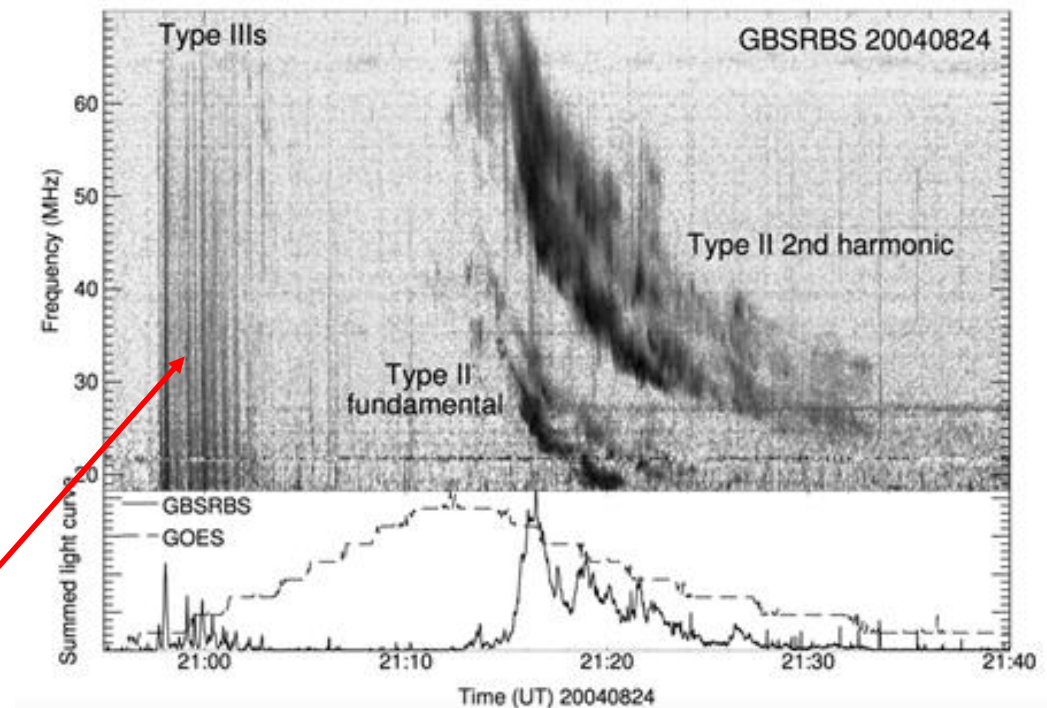
Bruit dans l'enregistrement  
Mots redondants

# Exemple de difficultés pour trouver l'information

- Emissions en rayons X et détection / caractérisation d'éruptions solaires



Autres signaux

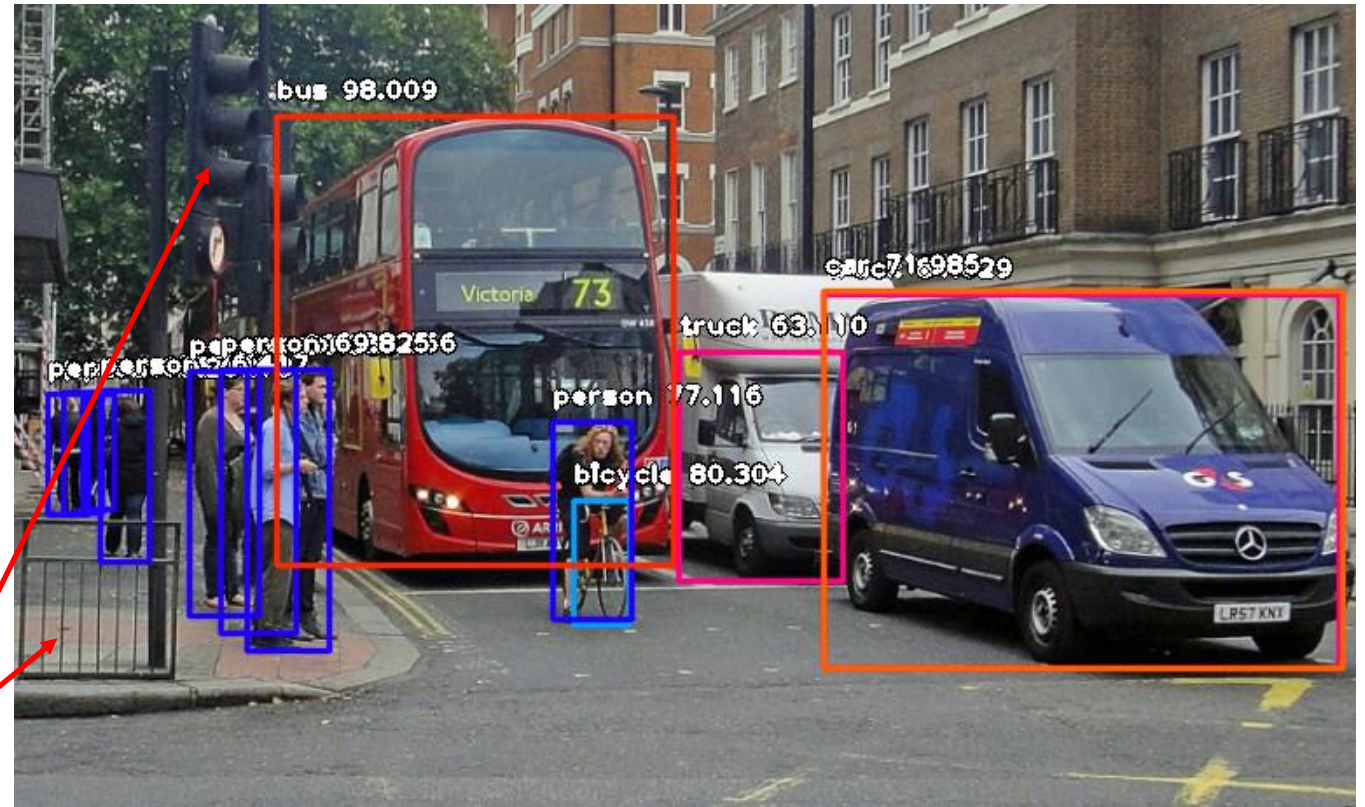


Bruit de mesure



# Exemple de difficultés pour trouver l'information

- Vidéo et détection / suivi de voitures et piétons

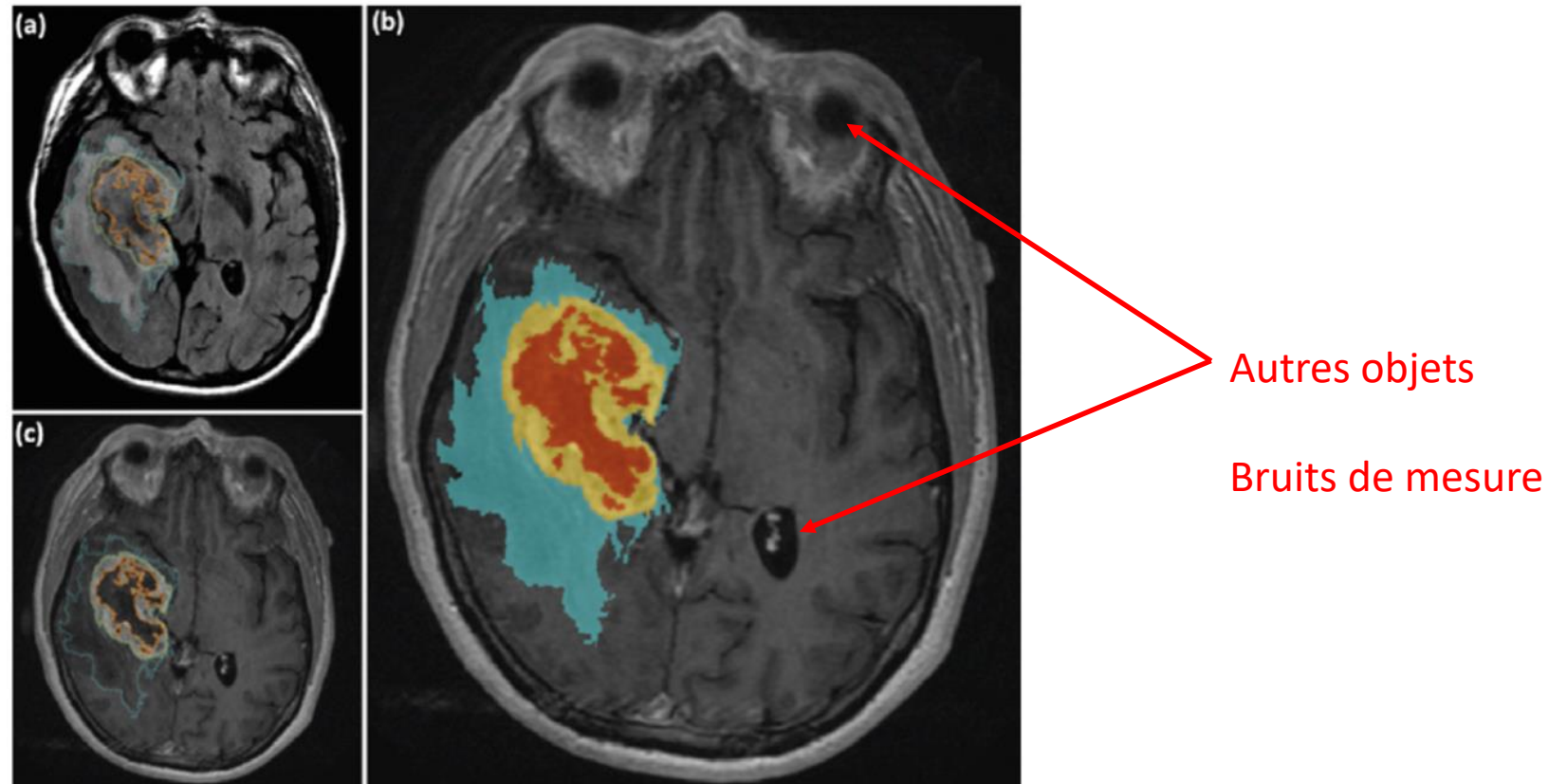


Autres objets

Les conditions météo peuvent changer

# Exemple de difficultés pour trouver l'information

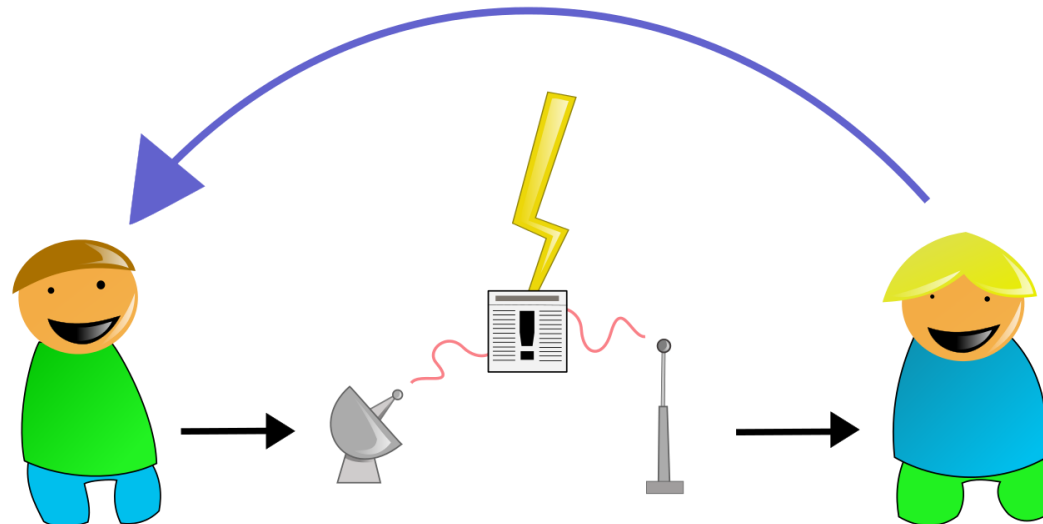
- IRM et détection / caractérisation de tumeurs



# Dégradations du message / des données

- Comment **recupérer** l'**information utile**  
des **données mesurées** (c.à.d. **transmises**)  
**pouvant être dégradée pendant la mesure ?**

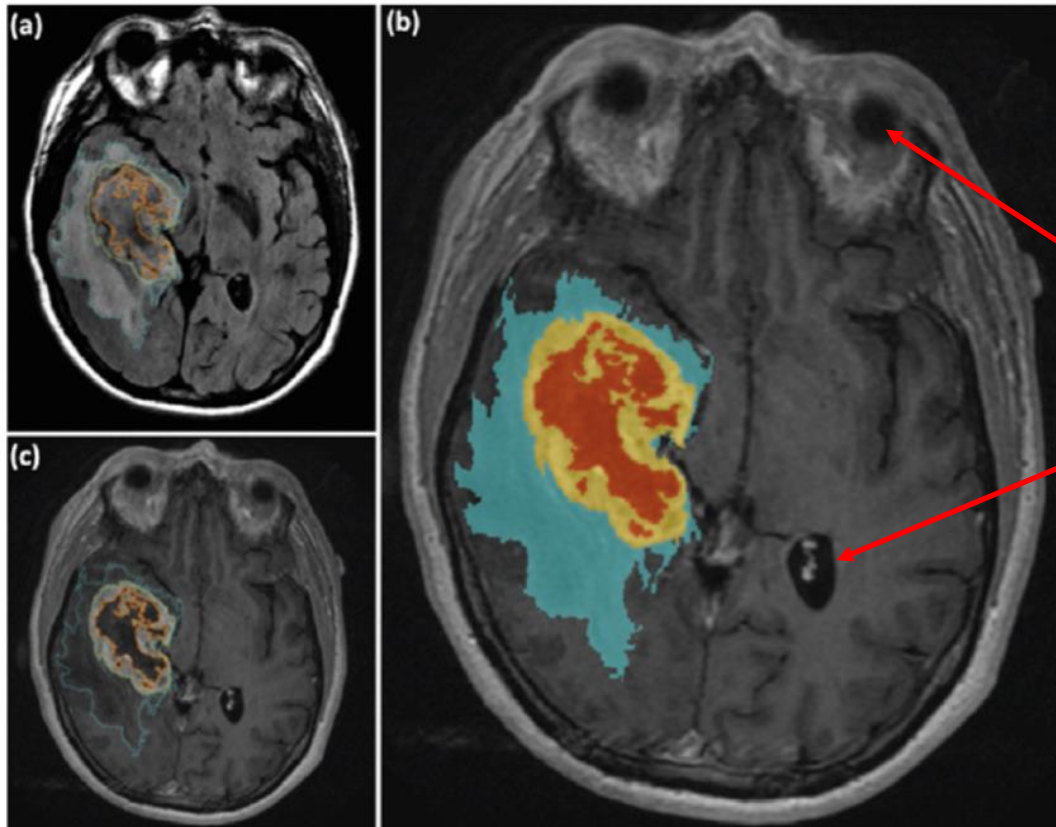
→ { Erreurs de mesure  
Autres objets et  
phénomènes capturés





# L'information dans son contexte

L'information est « utile » pour une application donnée



Objets intéressants mais inutiles pour suivre l'évolution de la tumeur

# L'information dans son contexte

Laquelle de ces deux phrases nous donne le plus d'information ?

- Hercule est un petit animal.
- Hercule est un poisson rouge.

# L'information dans son contexte

Laquelle de ces deux phrases nous donne le plus d'information ?

- Hercule est un petit animal.
- Hercule est un poisson rouge.

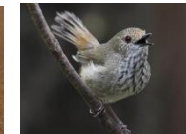
message + contexte

= véritable porteur d'information

Domaine des animaux



Domaine des petits animaux



# Questions à aborder

## Théorie de l'information en science des données

- Qu'est ce que l'information en science des données ?
- Quelles propriétés et défis pour cette information ?

## Quelques rappels et définitions fondamentales

- Probabilités
- Quantité d'information
- Entropie
- Indépendance
- Information mutuelle

Quelles applications en science des données ?

# Rappels de probabilité

- Variable aléatoire discrète  $X \in \mathcal{X}$ :

**Source** qui génère des objets (nombres, lettres, etc.) d'un ensemble  $\mathcal{X}$  en suivant une loi de probabilité donnée

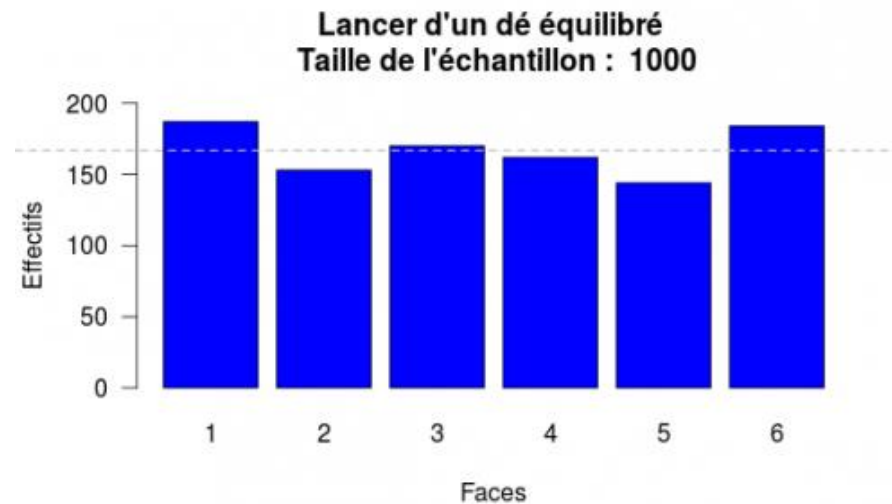
# Rappels de probabilité

- Variable aléatoire discrète  $X \in \mathcal{X}$  :

Source qui génère des objets (nombres, lettres, etc.) d'un ensemble  $\mathcal{X}$  en suivant une loi de probabilité donnée

Exemple :

Un dé équilibré génère des nombres entre 1 et 6, en suivant une loi uniforme



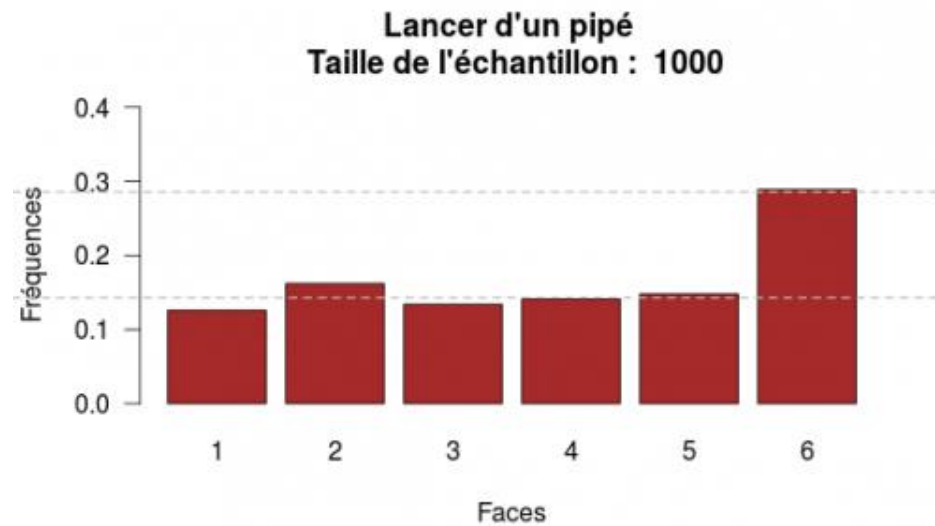
# Rappels de probabilité

- Variable aléatoire discrète  $X \in \mathcal{X}$  :

Source qui génère des objets (nombres, lettres, etc.) d'un ensemble  $\mathcal{X}$  en suivant une loi de probabilité donnée

## Exemple 2 :

Un dé **pipé** génère des nombres entre 1 et 6, en suivant une loi non uniforme



# Rappels de probabilité

- Variable aléatoire discrète  $X \in \mathcal{X}$  :

Source qui génère des objets (nombres, lettres, etc.) d'un ensemble  $\mathcal{X}$  en suivant une loi de probabilité donnée

## Exemple 3 :

Un texte est généré par une variable aléatoire qui génère des lettres de A à Z selon la distribution suivante :

<i>lettre</i>	<i>%</i>	<i>Lettre</i>	<i>%</i>	<i>Lettre</i>	<i>%</i>	<i>Lettre</i>	<i>%</i>	<i>Lettre</i>	<i>%</i>	<i>Lettre</i>	<i>%</i>
A	6,4	B	0,64	C	2,59	D	2,6	E	14,86	G	0,83
H	0,61	I	5,91	J	0,23	K	0,01	L	4,65	M	2,45
N	6,23	O	4,59	P	2,56	Q	0,81	R	5,55	S	6,97
T	5,72	U	5,06	W	0	X	0,31	Y	0,21	Z	0,08
Espace	18,35										
F	1,12	V	0,66								



# Rappels de probabilité

- **Valeurs** pouvant être générées par une variable aléatoire discrète  $X \in \mathcal{X}$  :

Les éléments de  $\mathcal{X}$

❑ Dé : {1; 2; 3; 4; 5; 6}

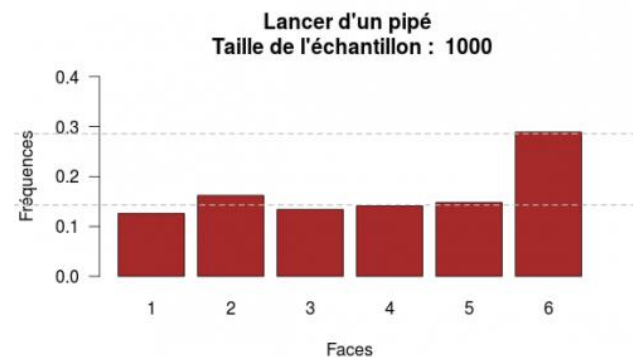
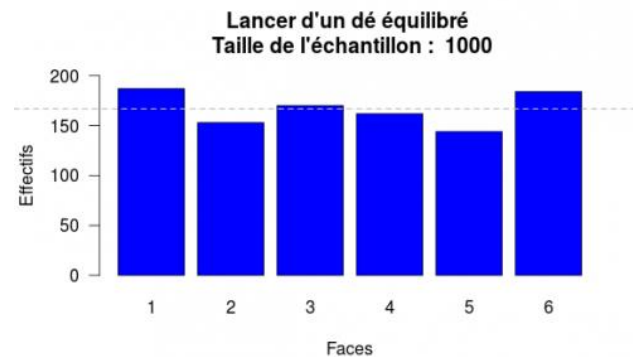
❑ Texte : {A; B; C; ....; Y; Z}

❑ etc.

# Rappels de probabilité

- Distribution / loi de probabilité :

Chance qu'une valeur donnée soit générée par la variable aléatoire



<i>lettre</i>	<i>%</i>	<i>Lettre</i>	<i>%</i>	<i>Lettre</i>	<i>%</i>	<i>Lettre</i>	<i>%</i>	<i>Lettre</i>	<i>%</i>	<i>Lettre</i>	<i>%</i>
A	6,4	B	0,64	C	2,59	D	2,6	E	14,86	G	0,83
H	0,61	I	5,91	J	0,23	K	0,01	L	4,65	M	2,45
N	6,23	O	4,59	P	2,56	Q	0,81	R	5,55	S	6,97
T	5,72	U	5,06	W	0	X	0,31	Y	0,21	Z	0,08
Espace	18,35										
F	1,12	V	0,66								

# Rappels de probabilité

- Distribution / loi de probabilité :

Probabilité de piocher une boîte bleue parmi toutes les boîtes ?

Probabilité de piocher une boîte grise parmi toutes les boîtes ?

...



# Rappels de probabilité

- Distribution / loi de probabilité :

Probabilité de piocher une boîte bleue parmi toutes les boîtes :  $\frac{n_{bleue}}{N}$

Probabilité de piocher une boîte grise parmi toutes les boîtes :  $\frac{n_{grise}}{N}$

...



# Rappels de probabilité

- Distribution / loi de probabilité :

Probabilité de la valeur bleue :  $\frac{n_{bleue}}{N}$

Probabilité de la valeur grise :  $\frac{n_{grise}}{N}$

...



# Rappels de probabilité

Questions ?

Comment mesure t'on l'information ?

