

# Recherche Zen

## Séance 6 : Pratiques éthiques

---

Carlos Ramisch and Manon Scholivet

Partly based on the course by Adeline Paiement

17 avril 2023

# Que fait-on de la vie ?

- Devenir / être chercheur.se
  - Autonomie, liberté thématique et méthodologique
  - Être son propre boss - charge supplémentaire
- Métier souvent associé à une passion, un engagement
  - **Frontière floue** pro/perso
  - Des projets en accord avec ses **valeurs** ?
  - Des activités qui **ont du sens** ?

## L'éthique

- En général : associée à la **morale, justice, principes**
- Ce cours : faculté de **se poser des questions** sur son travail
  - Bien être au travail
  - Limitations et biais
  - Changement social et environnemental
  - Aspects politiques, inégalités

Communication inter-personnelle au travail

Biais cognitif

Data management

Impact sociétal

Conditions de travail

- *Les cours, je gère. La recherche, ça me plait. Le plus compliqué, c'est les gens.*
- Améliorer sa communication = plus de sérénité au travail
- Communication écrite
  - Emails, notes, OdJ et CR de réunions
- Communication orale
  - Réunions, visios, prise de décisions

- Objet : quoi / qui / quand / où
- Première phrase : demande directe
  - Contextualiser après
  - Snippet de certains clients mail
- **Un email = une question/demande**
  - Mettre la question/demande **en gras**
  - Tout le monde **déteste** les emails longs
  - Apporter les précisions plus tard
  - Proposer une réunion si la discussion s'éternise/se complexifie

- **Heure d'envoi** - maximiser la probabilité de réponse
  - Programmer le message pour plus tard
- Utiliser des **dates complètes** (avec année et timezone)
  - *demain aprem* → *demain 13/04/2023 à 14 :00 CEST*
- Donner une **deadline** pour y répondre
- Prévoir la **relance** à cette date ("snooze")

# Organiser une réunion

- Préparer un **document de travail** partagé
  - Ordre du jour + compte-rendu
  - TODO-liste à la fin
- Indiquer systématiquement **l'heure de fin**
  - *réunion à 16h* → *réunion 16h-17h*
- Visio
  - Communication non verbale très limitée
  - Expliciter ce qui est évident
  - Structuration plus importante



- Documenter les **décisions** prises
  - Éviter les instructions incohérentes
- Rendre la vie du "moi du futur" plus zen
- Écrire = **organiser les idées**

# Conversations désagréables

- Il est toujours plus facile de ne rien dire
  - Zone de confort désagréable
- Long terme : conflits, stress, angoisse
  - Il vaut mieux le dire de travers que ne pas le dire
- Relations hiérarchiques – pas souvent nommées
- Astuce : s'entraîner sur des problèmes sans gros enjeu

Développé par le psychologue M. Rosenberg dans les années 60-70

1. **Description** : *Quand tu dis ...*

→ Focus sur le **comportement** pas la personnalité

→ Description objective est succincte

2. **Sentiment** : *Je me sens ...*

→ Parler en **première personne**

3. **Besoin** : *Or, j'ai besoin de me sentir ...*

→ Expliciter ses **besoins** affectifs

4. **Proposition** : *Est-ce qu'on pourrait ... ?*

→ Finir par une question

→ Garder un **contact visuel** et rester en **silence**

Communication inter-personnelle au travail

**Biais cognitif**

Data management

Impact sociétal

Conditions de travail

# Qu'est-ce qu'un biais ?

Un comportement pouvant induire des erreurs dans les résultats d'un procédé de recherche

# Qu'est-ce qu'un biais ?

Un comportement pouvant induire des erreurs dans les résultats d'un procédé de recherche

Par exemple :

- biais de confirmation
- biais de sélection
- biais de mesure

Biais nous faisant voir uniquement les résultats allant dans notre sens.

- “Pour le tagging, en moyenne, les résultats sur toutes les langues augmentent en utilisant le WALS ! C’est que ça marche”
- “Pour le parsing, en moyenne, les résultats sur toutes les langues baissent en utilisant le WALS.. Mais ce n’est pas grave, parce que pour cette langue particulièrement difficile, il y a une très légère augmentation. Donc ça marche ! ”

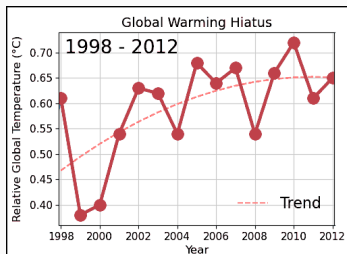
Comportement consistant à ne garder que les données qui vont dans le sens de l'hypothèse et à écarter les autres.

- 
1. Exemple issu de Wikipédia



# Cherry picking

Comportement consistant à ne garder que les données qui vont dans le sens de l'hypothèse et à écarter les autres.

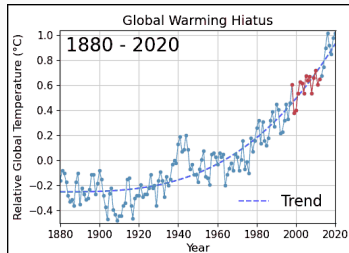
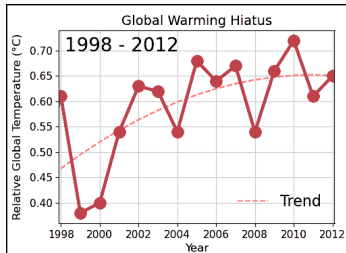


---

## 1. Exemple issu de Wikipédia

# Cherry picking

Comportement consistant à ne garder que les données qui vont dans le sens de l'hypothèse et à écarter les autres.



1

1. Exemple issu de Wikipédia

## Création d'une base de donnée non-représentative

- Étude de la durée de vie moyenne des humains
- Mais gros déséquilibre sur le nombre de femmes et d'hommes dans les données

## Mauvais choix de mesure

- Dataset de test : une centaine d'exemples positifs mais des milliers de négatifs
- Choix de l'accuracy/exactitude au lieu du f-score : fortement biaisé par le déséquilibre négatif/positif

La liste des biais est longue !

- Biais mnésique
- Biais de jugement
- Biais de raisonnement
- Biais liés à la personnalité

N'hésitez pas à consulter la liste !

[https://fr.wikipedia.org/wiki/Biais\\_cognitif](https://fr.wikipedia.org/wiki/Biais_cognitif)

Communication inter-personnelle au travail

Biais cognitif

Data management

Impact sociétal

Conditions de travail

# GDPR : general data protection regulation

- Concerns only **personal** data
  1. Inform contributors how the data will be used
  2. Provide access and possibility to correct data
  3. Allow data to be removed / forgotten
  4. Inform authorities of any data breach
  5. Ask permission for data use
- Gray zone :
  - De-identified user-generated text
  - Web-crawled corpora
  - Pre-trained language models
  - Web-published essays, novels, blogs, articles
    - Copyright

- **Temporary** : work in progress
  - Public git repo - refer to tags or commit numbers
  - Personal websites
  - Consistency can be challenging
  - Backup
- **Permanent** : data repository
  - Generic : Zenodo <https://zenodo.org/>
  - Specialised :
    - CLARIN-LINDAT <https://lindat.mff.cuni.cz/>
    - Ortolang <https://www.ortolang.fr/>
    - LDC <https://www ldc.upenn.edu/>
    - ELRA <http://catalog.elra.info/>



# FAIR principles

- **Findable**
  - unique identifier (DOI, Handle.net, URI ...)
  - meta-data
  - catalogue, search engine referencing
- **Accessible**
  - open protocols/formats for meta-data
  - authentication if needed
- **Interoperable**
  - well defined, standard, convenient format
  - use of FAIR vocabularies
- **Reusable**
  - clearly assigned licence
  - document sources
  - use widely adopted community standards

Source : <https://www.go-fair.org/fair-principles/>

- Describe meta-data in standard ways

## Composition

**What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?** Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.

The instances are movie reviews extracted from newsgroup postings, together with a sentiment polarity rating for whether the text corresponds to a review with a rating that is either strongly positive (high number of stars) or strongly negative (low number of stars). The sentiment polarity rating is binary {positive, negative}. An example instance is shown in figure 1.

**How many instances are there in total (of each type, if appropriate)?**

There are 1,400 instances in total in the original (v1.x versions) and 2,000 instances in total in v2.0 (from 2014).

**Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?** If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).

The dataset is a sample of instances. It is intended to be a random sample of movie reviews from newsgroup postings, with the

set of stars that was either given (based on the mapping from number of stars to polarity is given below in “Data Preprocessing”).

**Is there a label or target associated with each instance?** If so, please provide a description.

The label is the positive/negative sentiment polarity rating derived from the star rating, as described above.

**Is any information missing from individual instances?** If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

Everything is included. No data is missing.

**Are relationships between individual instances made explicit (e.g., users’ movie ratings, social network links)?** If so, please describe how these relationships are made explicit.

None explicitly, though the original newsgroup postings include poster name and email address, so some information (such as threads, replies, or posts by the same author) could be extracted if needed.

**Are there recommended data splits (e.g., training, development/validation, testing)?** If so, please provide a description of these splits, explaining the rationale behind them.

The instances come with a “cross-validation tag” to enable replication of cross-validation experiments; results are measured in classification accuracy.

<https://arxiv.org/abs/1803.09010>

- Open science, data sharing, reproducibility
- Creative Commons (CC-BY-4.0)  
<https://creativecommons.org/>
  - SA : share alike
  - NC : non commercial
  - ND : no derivatives
- Code : GNU GPL 2.0 <https://www.gnu.org/licenses/gpl>
  - Add LICENCE file to git repo/zip file
  - Add header to each code file

Communication inter-personnelle au travail

Biais cognitif

Data management

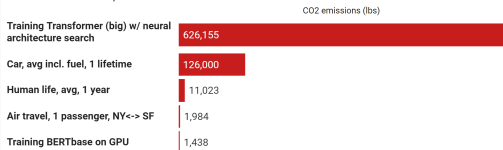
**Impact sociétal**

Conditions de travail

- *Energy and Policy Considerations for Deep Learning in NLP*
  - <https://aclanthology.org/P19-1355/>

## Carbon footprint comparison

Source: Strubell et al, 2019.



Reconstructed from: <http://arxiv.org/abs/1906.02243>

Model	Hardware	Power (W)	Hours	kWh-PUE	CO <sub>2</sub> e	Cloud compute cost
T2T <sub>base</sub>	P100x8	1415.78	12	27	26	\$41–\$140
T2T <sub>big</sub>	P100x8	1515.43	84	201	192	\$289–\$981
ELMo	P100x3	517.66	336	275	262	\$433–\$1472
BERT <sub>base</sub>	V100x64	12,041.51	79	1507	1438	\$3751–\$12,571
BERT <sub>base</sub>	TPUv2x16	—	96	—	—	\$2074–\$6912
NAS	P100x8	1515.43	274,120	656,347	626,155	\$942,973–\$3,201,722
NAS	TPUv2x1	—	32,623	—	—	\$44,055–\$146,848
GPT-2	TPUv3x32	—	168	—	—	\$12,902–\$43,008

Source: <https://aclanthology.org/P19-1355/>

## Some ethical considerations

- Too large models consume too many resources
  - <https://dl.acm.org/doi/10.1145/3442188.3445922>
- ML models learn gender bias, discrimination, etc.
  - <https://aclanthology.org/W19-3804/>
- "Debiasing" techniques do not work
  - <https://arxiv.org/abs/1903.03862>
- Predictive models can influence decisions

- ACM FAccT conference
  - <https://facctconference.org/>
- Domain-specific initiatives
  - Ethics in NLP : [https://aclweb.org/aclwiki/Ethics\\_in\\_NLP](https://aclweb.org/aclwiki/Ethics_in_NLP)
- Social networks, esp. twitter
  - Not always constructive



Example : ACL reproducibility guidelines

- Reproducibility
- Methodological soundness
- Ethical aspects
- Computational resources

Source: <https://aclrollingreview.org/responsibleNLPresearch/>

Communication inter-personnelle au travail

Biais cognitif

Data management

Impact sociétal

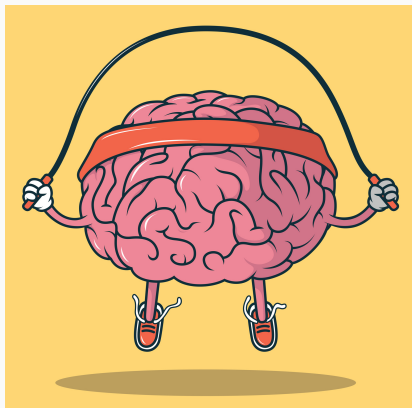
Conditions de travail

- Organisation des tâches
  - Trello, kanban, TODO liste
- Allouer plus de temps aux tâches
- Contraindre la "procrastination utile"
  - Enseignement, engagement collectif, etc.
- Équilibre entre enjeux et plaisir

- Tout est potentiellement intéressant
  - Mais qu'est-ce que ça **m'apporte** ?
  - **Pourquoi** je le fais ?
  - Reconnaissance, plaisir, rendre service, obligation, ennui . . .
- Dire "non"
  - Template email, calendrier, "non" temporaire

Quel est notre principal outil de travail ?

Quel est notre principal outil de travail ? Le cerveau !



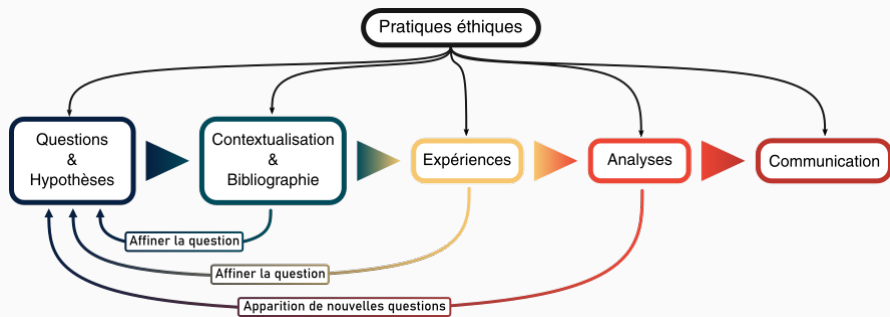
## Risques du métier : maltraiter son cerveau

- Burnout
  - Irritabilité, manque de motivation, fatigue
- Attaque de panique
  - Paralysie, douleur poitrine, peur de mourir
- Angoisse
  - Stress, souffle, tensions
- Dépression
  - Fatigue, immobilité, tristesse, manque de motivation

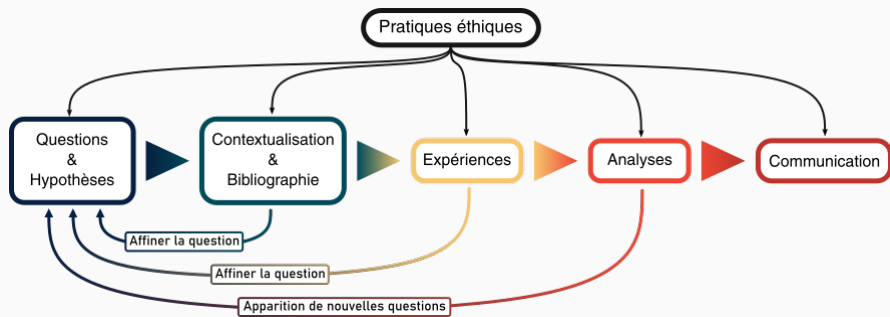
- Partager avec les collègues
- Se reposer, prendre des vacances
- Hygiène de vie : p.ex. pas de mail le soir
- Psychothérapie
- Sport
- Lâcher prise : rien n'est très grave



# Rappel : un idéal



# Rappel : un idéal



But : **faire de notre mieux** pour contribuer à la science  
Sans devenir (trop) fou, voire en restant **zen**