Recherche Zen
Séance 6 : Pratiques éthiques

Carlos Ramisch and Manon Scholivet
Partly based on the course by Adeline Paiement

22 novembre 2023

Que fait-on dans 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 ?

Disclaimer

L'éthique

- En général : associée à la morale, justice, principes
- Ce cours : faculté de se poser des questions sur son travail
  - $\longrightarrow$ Reproductibilité et réplicabilité
  - $\longrightarrow$ Limitations et biais
  - $\longrightarrow$ Impact social et environnemental
  - $\longrightarrow$ Communication et bien être au travail

Plan

The devil is in the details

Some details may have great impact on conclusions

- Experimental conditions
- Hyperparameters
- Overfitting
- Model instability
- Reproducibility, replicability
- ...

Experimental conditions

Experimental conditions can influence conclusions

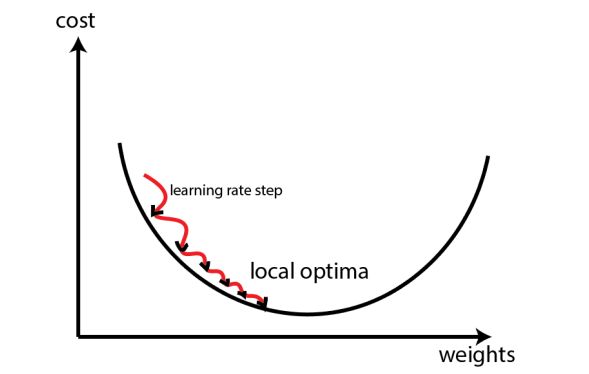
→ Only compare what is comparable

- Amount of supervision
  - → Supervised, unsupervised, semi-supervised
  - → Zero-shot, one-shot, few-shot, ...
- Ablation studies
  - → What part(s) of my model influence which results ?

Hyperparameters

Common hyperparameters in neural models (1/2)

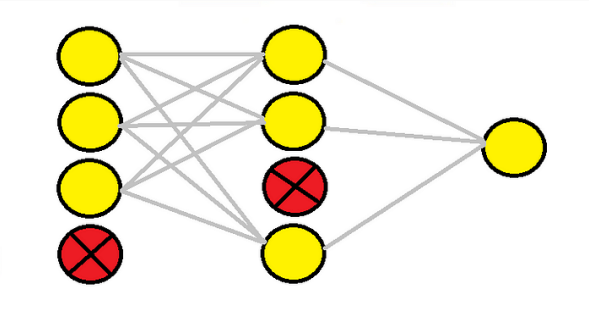
- Learning rate
  - → Speed at which parameters are updated
- Learning strategy
  - → Adam, SGD, warmup steps,...
- Number of epochs
  - → Iterations over full dataset



Hyperparameters

Common hyperparameters in neural models (2/2)

- Batch size
  - → Fast processing vs. fast convergence
- Dropout ratios
  - → Prevent memorisation
- Model capacity
  - → Layer dimensions, embedding dimensions
  - → Number of stacked layers, attention heads



Wooclap

Wooclap time !

Hyperparameter search

Best hyperparameter values ? Hyperparameter tuning !

- Grid search
  - → All possible (discretised) value combinations
- More sophisticated strategies
  - → Bayesian search
  - → Random search
- Specialised libraries
  - → Raytune, optuna, . . .

Hyperparameter search

Best hyperparameter values ? Hyperparameter tuning !

- Grid search
  - → All possible (discretised) value combinations
- More sophisticated strategies
  - → Bayesian search
  - → Random search
- Specialised libraries
  - → Raytune, optuna, ...

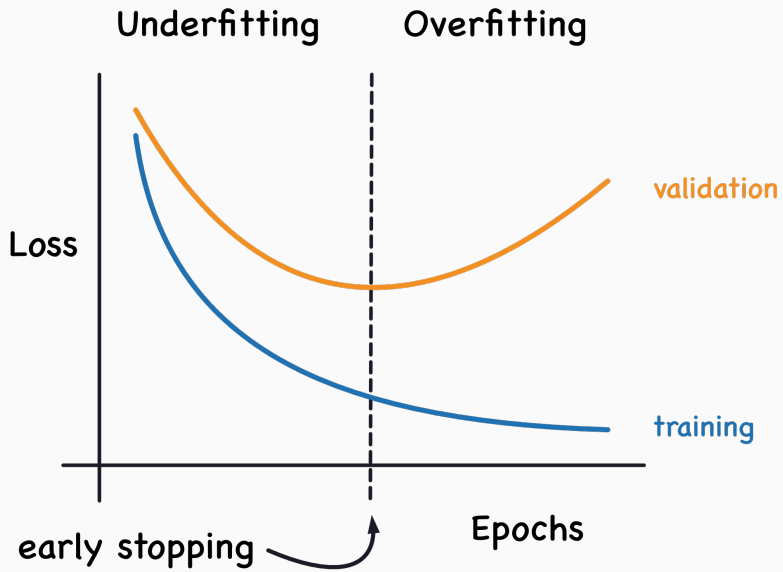
Unavoidable but usually not very interesting

Wooclap

Wooclap time !

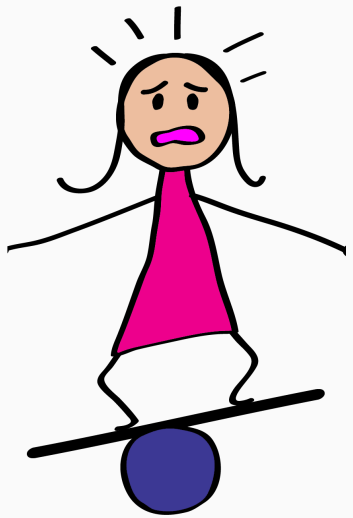
Overfitting

- The model "overfits" if it memorises the training set
- Rule of thumb of pre-neural models :
  - $\rightarrow$ Less features than data items
- Learning curves on dev set
- Early stopping based in dev set performance



Model instability

- Same hyperparameters, but different random seeds
  - Parameter initialisation
  - Order of inputs/batches
- Substantially different results
  - Some data orders/initializations consistently better



Assessing and preventing instability

- Early stopping may be effective
- Report averages, error bars, confidence intervals
  - Re-run several times with different orders/random seeds
  - Explicitly set `random.seed` (for each lib)
  - Record and publish random seed values
- How to analyse the results of several runs ?

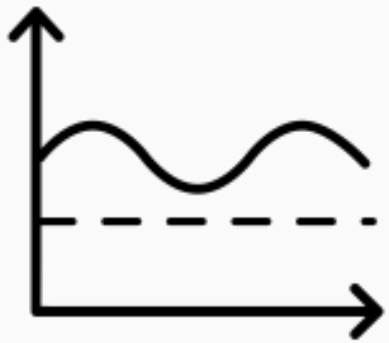
  $\rightarrow$ Majority vote among predictions

  $\rightarrow$ Ensemble models may improve results !
- Further reading : `https://arxiv.org/abs/2002.06305`

Baseline

A model is never good or bad per se

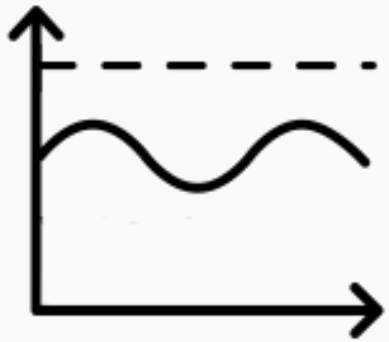
- Situate the model performance wrt. a simpler model
  - → Baseline – simple model for the task
- Examples
  - → Random prediction (too simple ?)
  - → Majoritary class
  - → A good model 5 years ago
  - → An interpretable model (rules, thresholds)
  - → State-of-the-art model published last month (too complex ?)



Topline

A model is never good or bad per se

- Situate the model performance wrt. a better model
  - → Topline – upper bound for the performance
- Examples
  - → State-of-the-art model published last month
  - → Large model released by big tech company
  - → Human annotator performance/agreement
  - → Same experiment in unrealistic (easy) condition

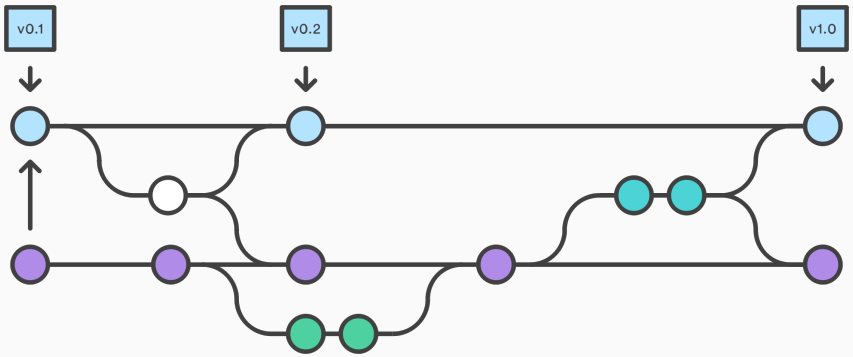


Experiments management

- Logbook
  - Experimental conditions for each result
  - Raw results and links to results
  - Write down ideas, hypotheses, etc.
- Experiments management platform
  - Tensorboard, RayTune, MLFlow, Lightning

Experiments management

- Parallelisation : clusters
  - → Distributed supercomputer to run many experiments simultaneously
  - → National : Jean Zay, AMU : mesocentre, local...
- Git : branches, merge requests, CI for testing
- Overleaf : collaborative LaTeX writing



Reproducibility vs. replicability

- Results are reproducible
  - Data available under open licences
  - Model/code shared under open licences
  - Parameters and hyperparameters described
  - Computational requirements reasonable
- Results are replicable
  - Robust to other datasets
  - Robust to different experimental conditions
  - Robust across conditions

Source: `https://acl-reproducibility-tutorial.github.io/`

Wooclap

Wooclap time !

Plan

Personal data

- Anonymisation
  - Remove all information which allows identifying individuals
  - Aggregate, shuffle
- Pseudo-anonymisation or de-identification
  - Remove identity-related information (name, phone, email)
  - Analysis/crossing could recover individuals identities
- In practice : complete anonymisation barely impossible

GDPR : general data protection regulation

- Concerns only personal data
- GDPR in a nutshell :
  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

GDPR gray zones

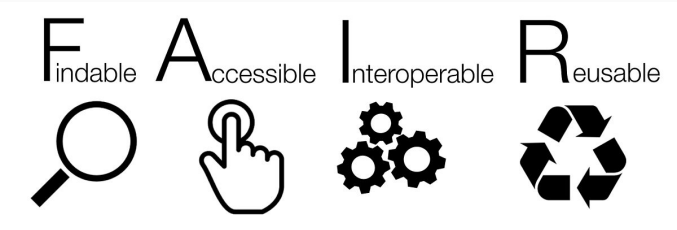
- De-identified user-generated content
- Web-crawled content
- Models pre-trained on personal data
- Web-published artwork (essays, novels, blogs, articles)
    - Copyright vs. GDPR ?

Data repositories

- Temporary : work in progress
  - Public git repo - refer to tags or commit numbers
  - Personal website
  - Consistency can be challenging
  - Backup is important
- Permanent data repository
  - Generic : Zenodo `https://zenodo.org/`
    → Safe, permanent, citable (DOI)
  - Specialised, e.g. for linguistic datasets :
    → CLARIN-LINDAT, Ortolang, LDC, ELRA, ...

FAIR principles

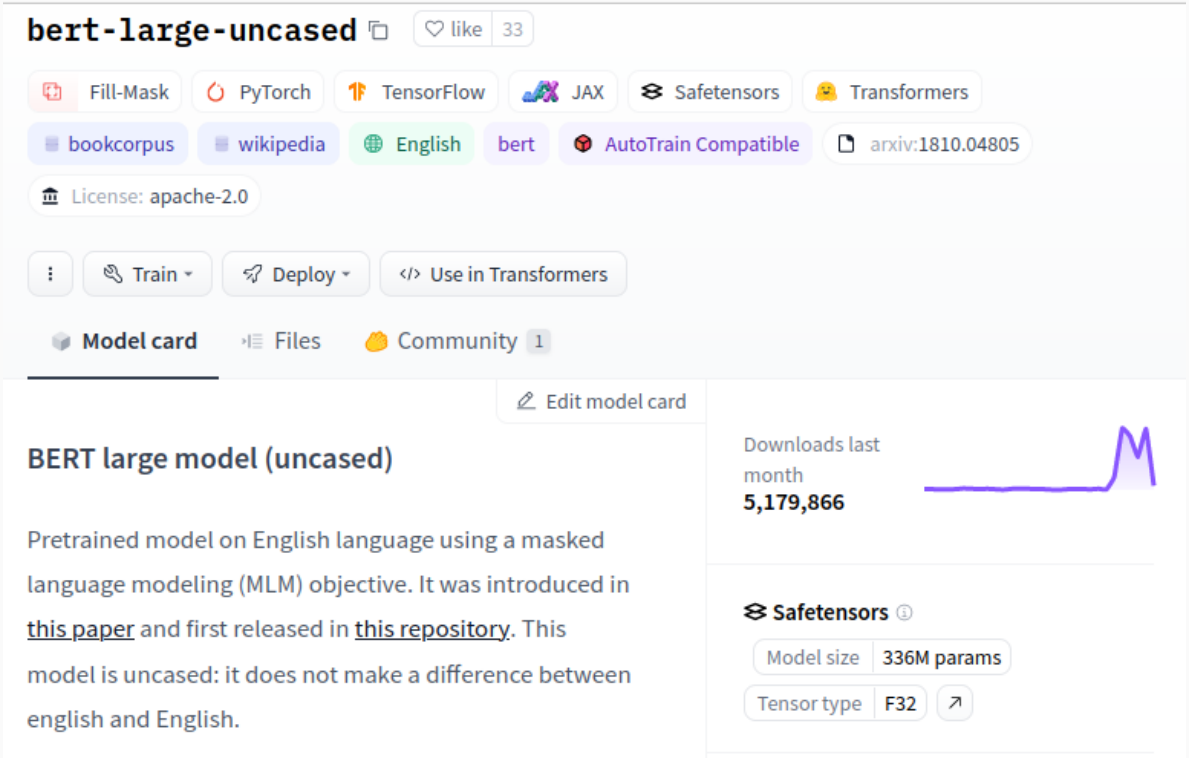
- Findable
  - unique identifier (DOI, Handle.net, URI ...)
  - present in catalogues and search engines
- Accessible
  - open protocols/formats for meta-data
- Interoperable
  - well defined, standard, convenient format
- Reusable
  - clearly assigned licence, document sources
  - use widely adopted community standards



Source : `https://www.go-fair.org/fair-principles/`

Data/model sheets

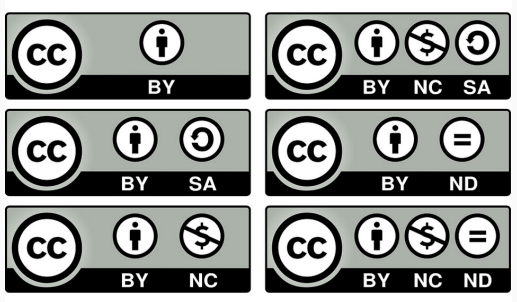
- Describe meta-data in standard ways (e.g. Huggingface)



- Further reading : `https://arxiv.org/abs/1803.09010`

Licences

- Open science, data sharing, reproducibility
- Data : creative commons 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



Plan

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 cognitif

- Schéma systématique de déviation du jugement, par rapport à la norme ou à la rationalité
- Perception biaisée de l'entrée $\rightarrow$ "réalité subjective"
- Peut conduire à
  - Perception distordue
  - Jugement inexact
  - Interprétation illogique
  - Irrationalité
  - ...

Source: `https://en.wikipedia.org/wiki/Cognitive_bias`

Biais de confirmation

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 ! ”

Biais de confirmation

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 ! "

- Regarder des tendances $\rightarrow$ biaisé par nos croyances

Wooclap

Wooclap time !

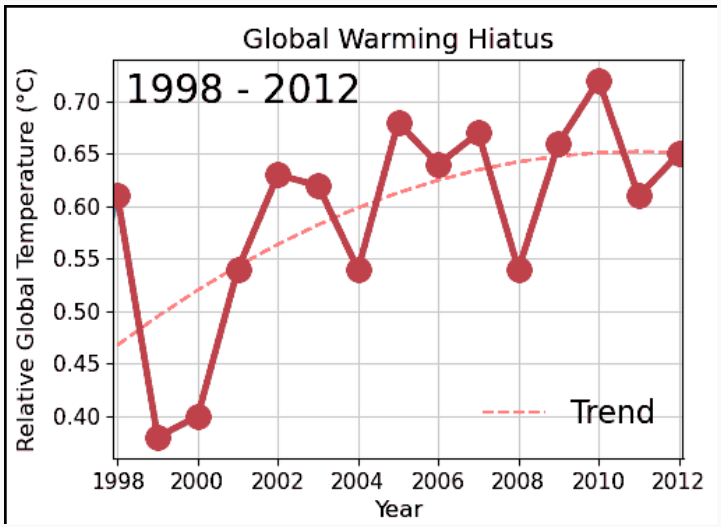
Cherry picking [1]

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

1. L'art d'exclure ou picorage

Cherry picking [1]

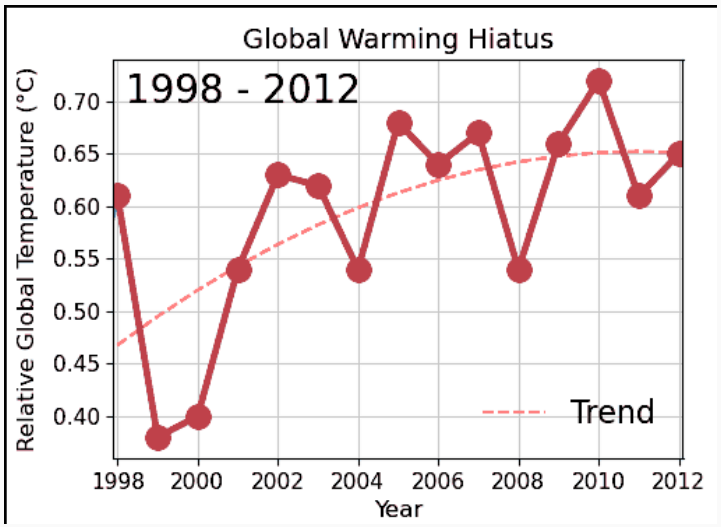
Comportement consistant à ne garder que les données qui vont dans le sens de l'hypothèse et à écarter les autres.

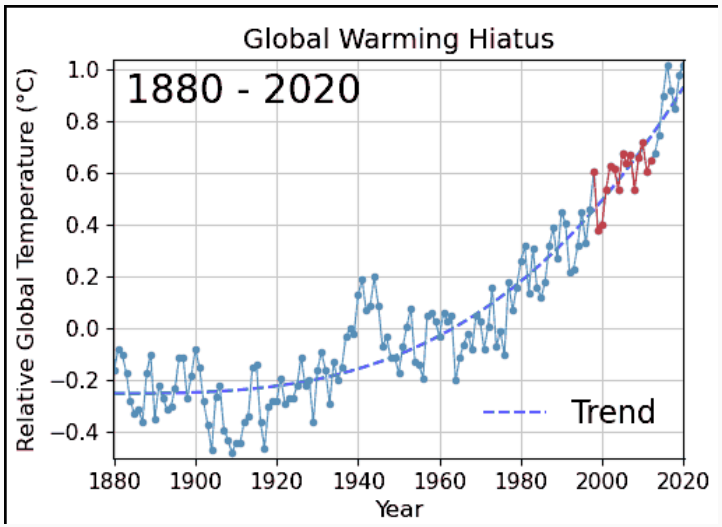


1. L'art d'exclure ou picorage

Cherry picking [1]

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





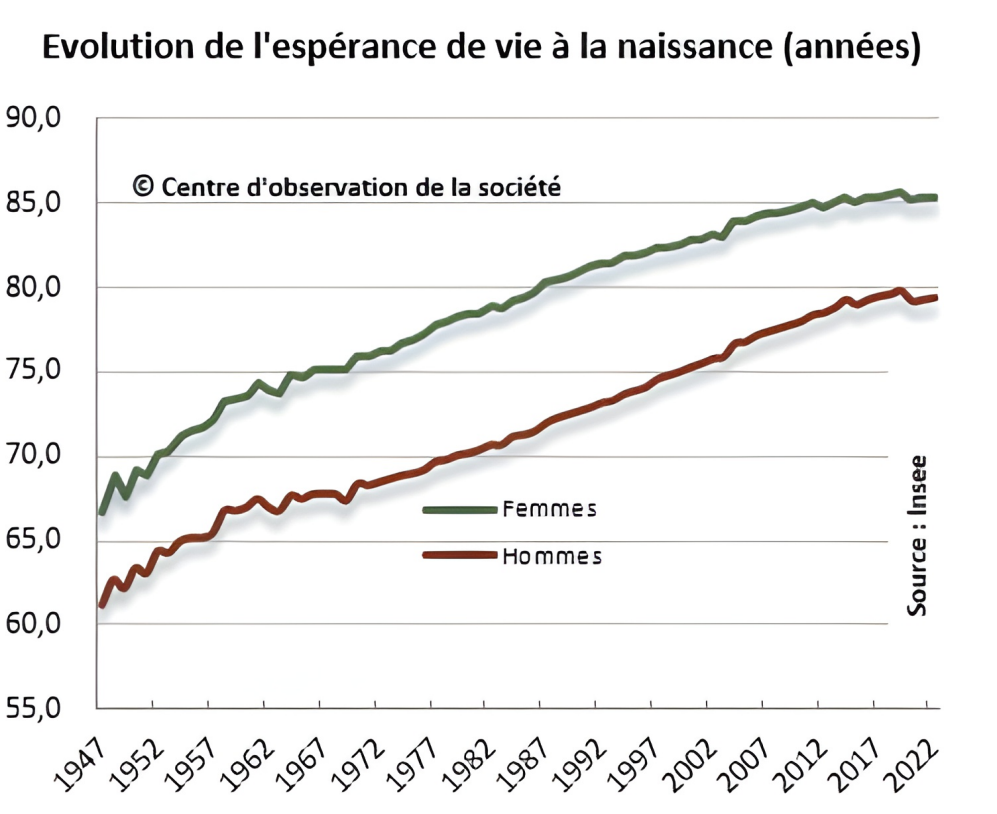
Exemple issu de Wikipédia

1. L'art d'exclure ou picorage

Biais de sélection

Création d'une base de donnée non-représentative

- Étude de la durée de vie moyenne des humains
- Oups ! 90% du dataset est constitué d'hommes



Biais de mesure

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
- Choisir les métriques en amont, pas reporter uniquement les métriques qui nous intéressent à posteriori

Biais de conformité

- Biais d'appartenance à un groupe
  - $\rightarrow$ Faire ce que tout le monde a toujours fait dans la communauté
- Exemple : utiliser des benchmarks ou métriques problématiques (p.ex. BLEU pour la traduction automatique) pour pouvoir se comparer à l'état de l'art

Toujours plus de biais

La liste des biais est longue !

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

Pour lutter contre les biais, il faut les connaître !

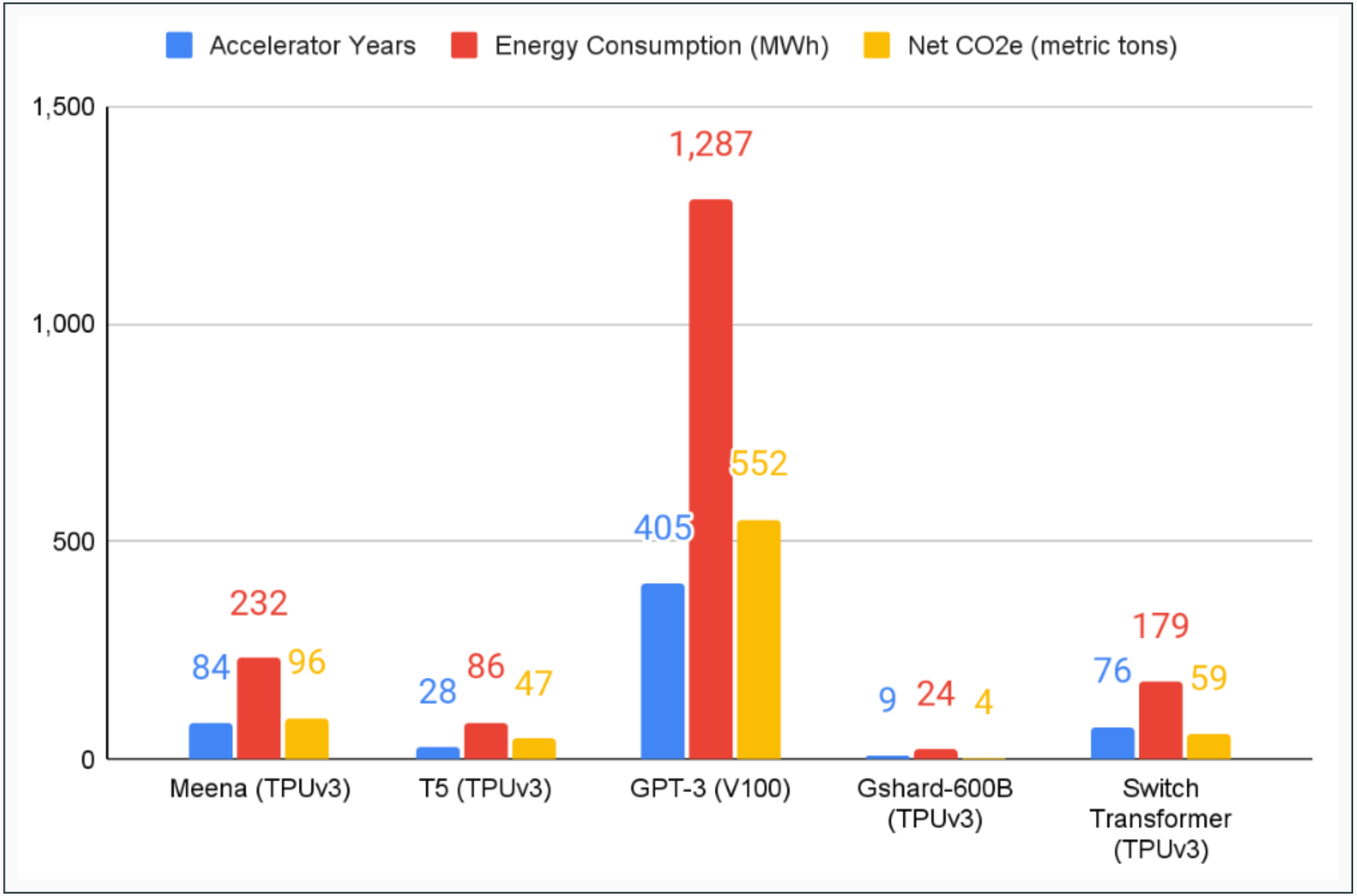
N'hésitez pas à consulter la liste :
`https://fr.wikipedia.org/wiki/Biais_cognitif`

Connaître et communiquer sur les biais

- Dans certaines conférences, section “Limitations” obligatoire dans les articles
  - $\longrightarrow$ On peut/doit lister les biais identifiés
  - $\longrightarrow$ On n’est pas obligés d’avoir des solutions

Plan

Energy consumption

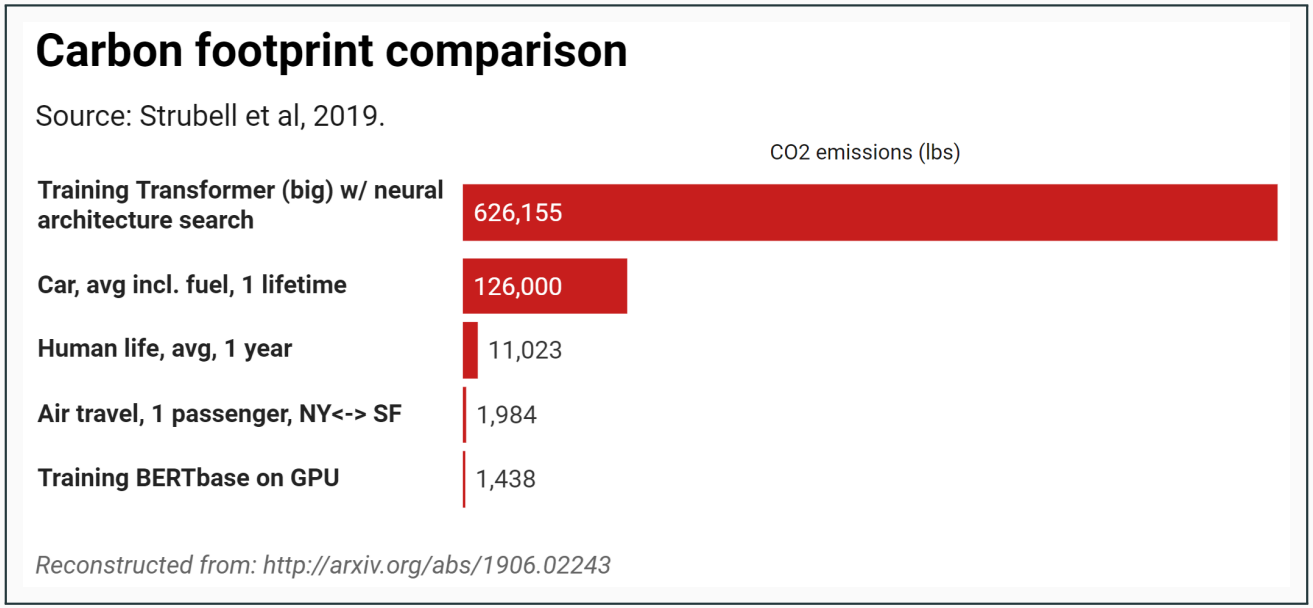


Source: https://arxiv.org/abs/2104.10350

Wooclap

Wooclap time !

Carbon footprint



- Average French person : 12 tCO2 / year
- Training transformer LM : 312.6 tCO2

Source: `https://aclanthology.org/P19-1355/`

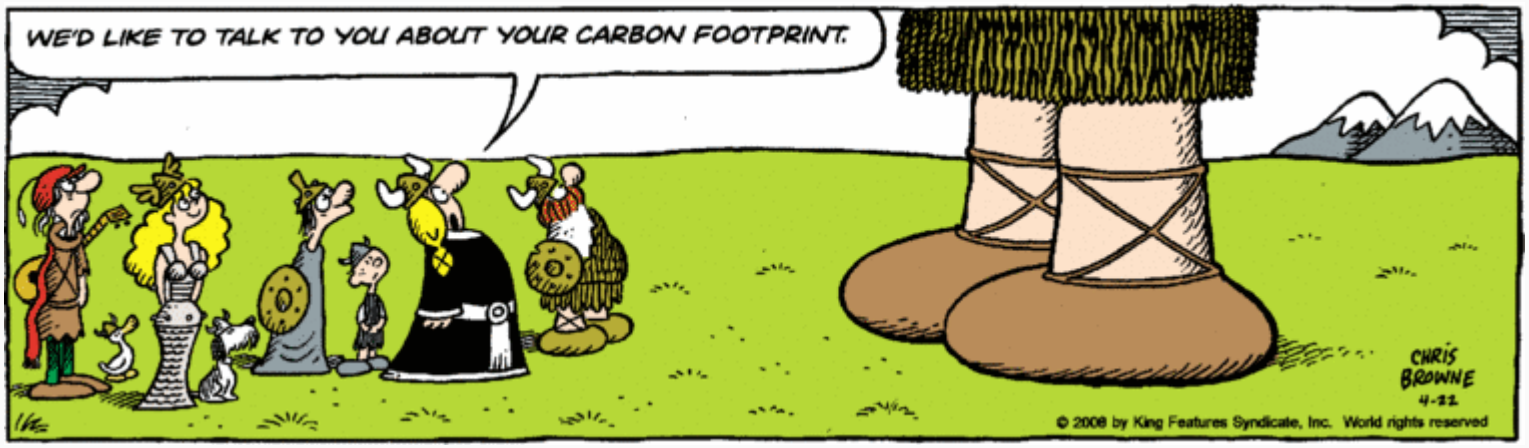
Measuring carbon footprint

- Specialised tools

  → Code Carbon, . . .

- Measuring carbon footprint is hard

  → `https://aclanthology.org/2021.sustainlp-1.2/`



Accessibility

Model	Hardware	Power (W)	Hours	kWh·PUE	CO_2e	Cloud compute cost
$T2T_{base}$	P100x8	1415.78	12	27	26	$41–$140
$T2T_{big}$	P100x8	1515.43	84	201	192	$289–$981
ELMo	P100x3	517.66	336	275	262	$433–$1472
$BERT_{base}$	V100x64	12,041.51	79	1507	1438	$3751–$12,571
$BERT_{base}$	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

- According to OpenAI, running chatGPT costs 700K US$ / day
  - → Most costs go into server maintenance and electricity
- Public research budgets not competitive

Source: `https://aclanthology.org/P19-1355/`

Biases in ML models

- Problems in the data used to train the models
  - → Lack of diversity
  - → Contain harmful content
  - → Reflect implicit stereotypes
- ML models are biased against historically disadvantaged groups
  - → E.g. gender bias in NLP : `https://aclanthology.org/W19-3804/`
- "Debiasing" techniques are of limited impact
  - E.g. gender bias in NLP : `https://arxiv.org/abs/1903.03862`
- Predictive models can be more biased than training data
  - → E.g. bias amplification : `https://aclanthology.org/D17-1323/`

Going further

- ACM FAccT conference `https://facctconference.org/`
- Domain-specific resources
  - → Ethics in NLP : `https://aclweb.org/aclwiki/Ethics_in_NLP`
  - → NLP bias survey : `https://aclanthology.org/2020.acl-main.485/`
  - → E.g. `https://members.loria.fr/KFort/teaching/summer-schools/`
- Social networks, especially Twitter
  - → Not always constructive

Responsible research

Example : ACL guidelines for responsible research

- Reproducibility
- Methodological soundness
- Ethical aspects
- Computational resources

Source: `https://aclrollingreview.org/responsibleNLPresearch/`

Plan

Communication inter-personnelle au travail

- *Les cours, je gère. La recherche, ça me plaît. Le plus compliqué, c'est les gens.*
- Améliorer sa communication = plus de sérénité au travail
- Communication écrite

  $\rightarrow$ Emails, notes, ordre du jour et compte-rendu de réunions
- Communication orale

  $\rightarrow$ Réunions, visios, prise de décisions

Écrire un email

- 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

Avoir une réponse

- 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")

Wooclap

Wooclap time !

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

Prise de notes

- Documenter les décisions prises
  - $\rightarrow$ 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

Communication non violente

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

Plan

Gestion du temps

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

Définir ses priorités

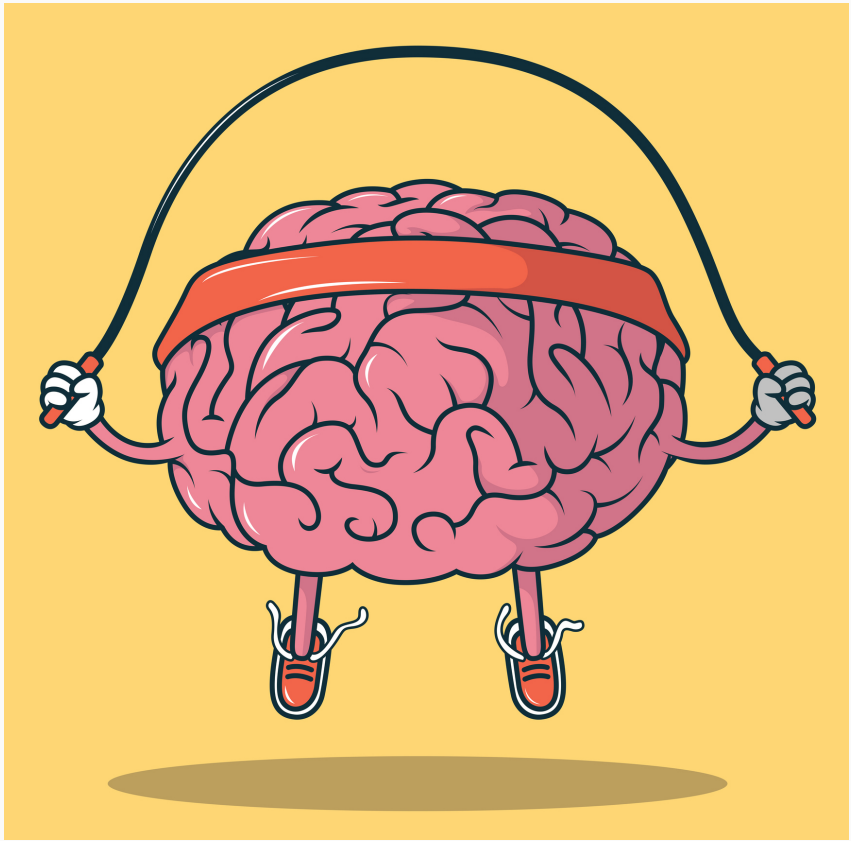
- Tout est potentiellement intéressant
  - → Mais qu'est-ce que ça m'apporte ?
  - → Pourquoi je le fais ?
  - → Reconnaissance, plaisir, rendre service, obligation, ennui ...
- Apprendre à dire "non"
  - → Template email, calendrier, "non" temporaire

Santé mentale

Quel est notre principal outil de travail ?

Santé mentale

Quel est notre principal outil de travail ? Le cerveau !



Stress au travail

Source: Adapté depuis `https://www.inrs.fr/media.html?refINRS=A%20789`

Santé mentale

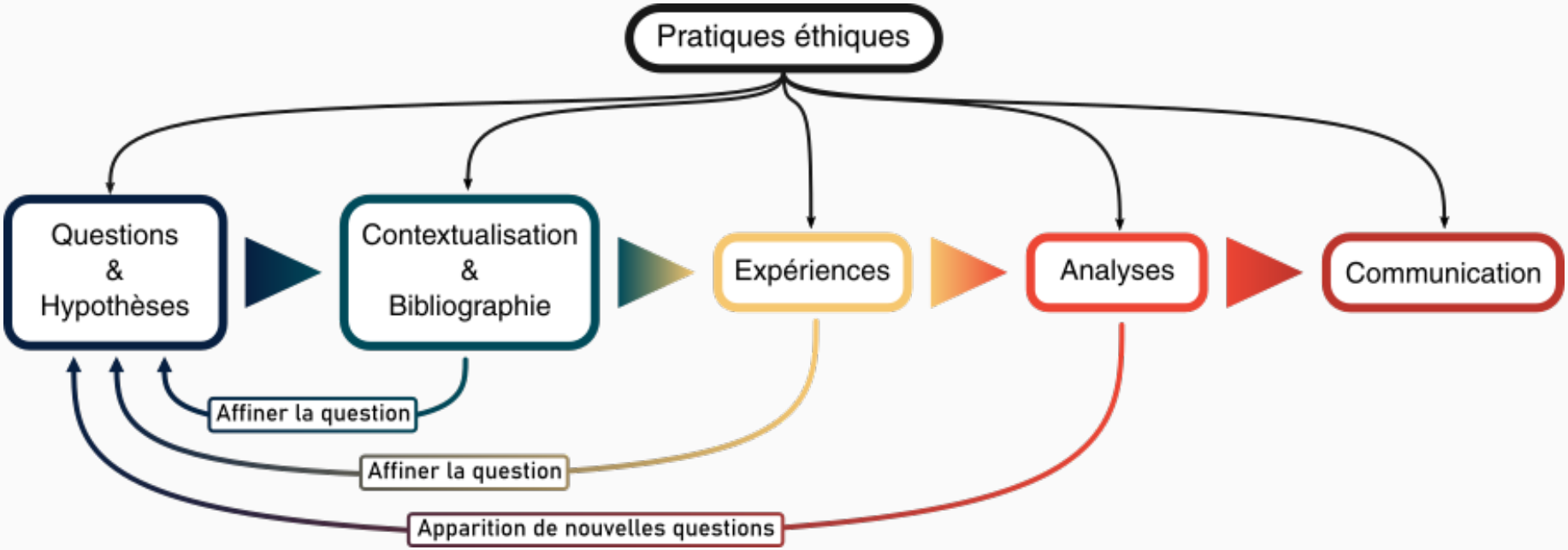
Risques psycho-sociaux : maltraiter son cerveau

- Burnout
  → Irritabilité, manque de motivation, fatigue
- Angoisse
  → Stress, souffle irrégulier, attaque de panique
- Dépression
  → Fatigue, immobilité, tristesse, manque de motivation
- Addictions
  → Perte de contrôle, craving, répercussions

Quelques idées

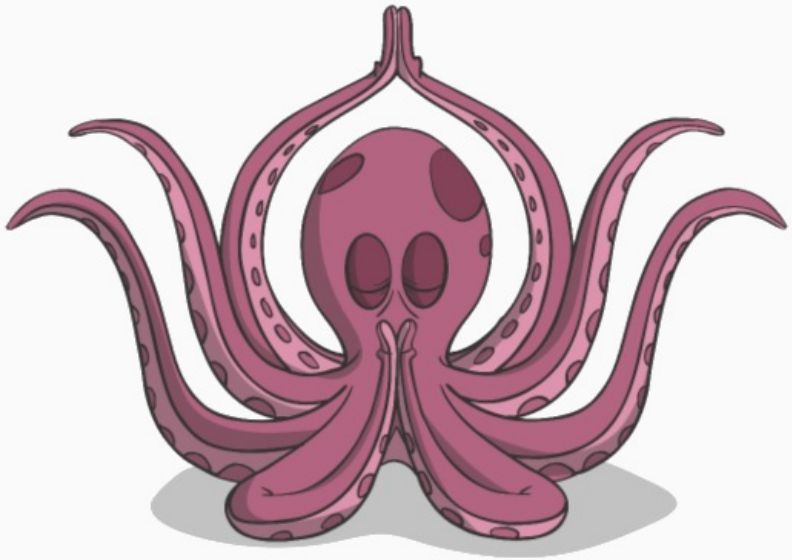
- Partager avec les collègues
- Se reposer, prendre des vacances
- Hygiène de vie : p.ex. pas de mail le soir
- Psychothérapie
- Sport
- Lâcher prise : rien n'est très grave

Rappel : un idéal



La recherche zen : un idéal

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



Sources

- Cours d'Adeline Paiement
- Wikipedia
- Google images
- Cours de Karen Fort :

  `https://members.loria.fr/KFort/teaching/summer-schools/`