

Développement logiciel : Reproductibilité

Juliette Luiselli
Équipe Beagle, INRIA, Lyon & INSA Lyon
juliette.luiselli@inria.fr

7th April 2025
4 BIM

Introduction

Est-ce que la Terre est plate ?

Introduction

Est-ce que la Terre est plate ?

Est-ce que l'homéopathie fonctionne ?

Introduction

Est-ce que la Terre est plate ?

Est-ce que l'homéopathie fonctionne ?
Est-ce que l'hydroxychloroquine soigne le covid ?

Introduction

Est-ce que la Terre est plate ?

Est-ce que l'homéopathie fonctionne ?

Est-ce que l'hydroxychloroquine soigne le covid ?

Est-ce que les vaccins causent l'autisme ?

Introduction

Est-ce que la Terre est plate ?

Est-ce que l'homéopathie fonctionne ?

Est-ce que l'hydroxychloroquine soigne le covid ?

Est-ce que les vaccins causent l'autisme ?

Est-ce que le froid rend malade ?

Introduction

Est-ce que la Terre est plate ?

Est-ce que l'homéopathie fonctionne ?

Est-ce que l'hydroxychloroquine soigne le covid ?

Est-ce que les vaccins causent l'autisme ?

Est-ce que le froid rend malade ?

Comment le savez-vous ? Quelles en sont les preuves ?

LES NIVEAUX DE PREUVE

-> Une preuve est un élément susceptible d'établir la **réalité** d'un fait ou d'une proposition.
-> Les témoignages, expériences personnelles, études scientifiques, anecdotes et autres éléments servant à appuyer un propos ne **bénéficient pas tous de la même fiabilité**. Il existe une hiérarchisation, établie en fonction du niveau de preuve apporté. En voici un récapitulatif.

<ul style="list-style-type: none">- CONSENSUS SCIENTIFIQUE- MÉTA-ANALYSES D'ESSAIS COMPARATIFS RANDOMISÉS		NIVEAU DE PREUVE ÉLEVÉ
<ul style="list-style-type: none">- ESSAIS COMPARATIFS RANDOMISÉS DE FAIBLE PUISSANCE- ÉTUDES DE COHORTE		PRÉSUMPTION SCIENTIFIQUE
<ul style="list-style-type: none">- ÉTUDES CAS-TÉMOINS- ÉTUDES COMPARATIVES COMPORTANT DES BIAS IMPORTANTS- ÉTUDES RÉTROSPECTIVES- SÉRIE DE CAS		NIVEAU DE PREUVE FAIBLE
<ul style="list-style-type: none">- PAROLE D'EXPERT·E (RECONNU·E DANS SON DOMAINE D'EXPERTISE)		ÉLÉMENTS NE RELEVANT PAS DE LA PREUVE
<ul style="list-style-type: none">- ANECDOTE PERSONNELLE, TÉMOIGNAGE RAPPORTÉ		
<ul style="list-style-type: none">- RUMEUR, "BON SENS", SAGESSE POPULAIRE, CROYANCE, OPINION		

Remarques complémentaires :

- > Ce qui est affirmé sans preuve peut être rejeté sans preuve ;
- > Une information délivrée sans preuve peut toutefois être juste ;
- > L'absence de preuve ne permet pas de prouver l'existence ou l'inexistence d'une chose
- > Les affirmations extraordinaires nécessitent des preuves extraordinaires (maxime de Hume)
- > La science ne prétend pas délivrer de vérité absolue, mais plutôt un *état de l'art* correspondant à ce qu'il est raisonnable de penser en fonction des données disponibles à un instant donné ;
- > Une démarche rationnelle rigoureuse consiste à tirer des conclusions en fonction des faits, et non à chercher des faits pouvant aller dans le sens d'une conclusion établie à l'avance.



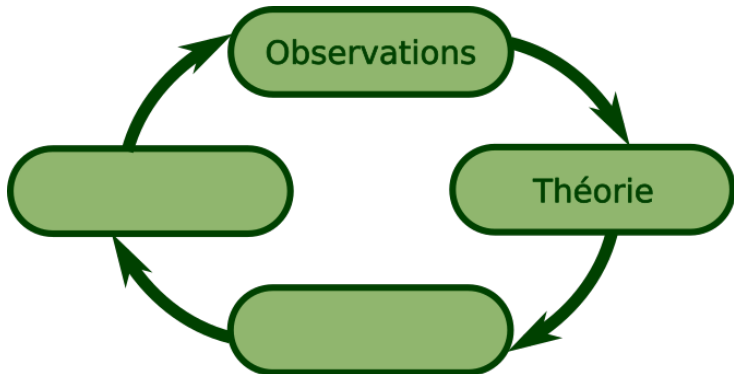
Comment sait-on ce qu'on sait – échelle individuelle et échelle collective ?

Contents

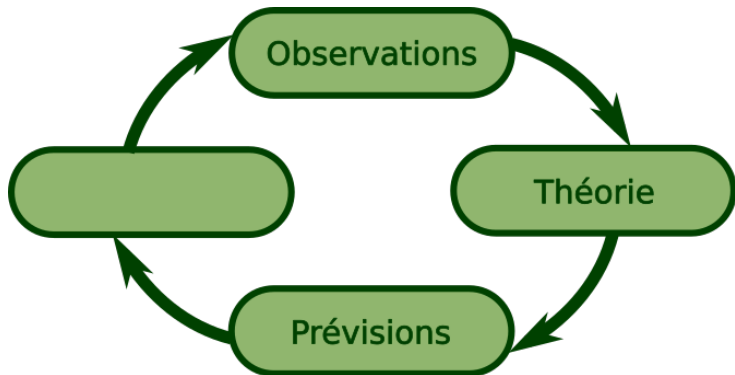
- 1 Introduction
- 2 La méthode scientifique
- 3 Reproductibilité, répétabilité et réplicabilité
- 4 Conseils pratiques
- 5 Conclusion

La méthode scientifique

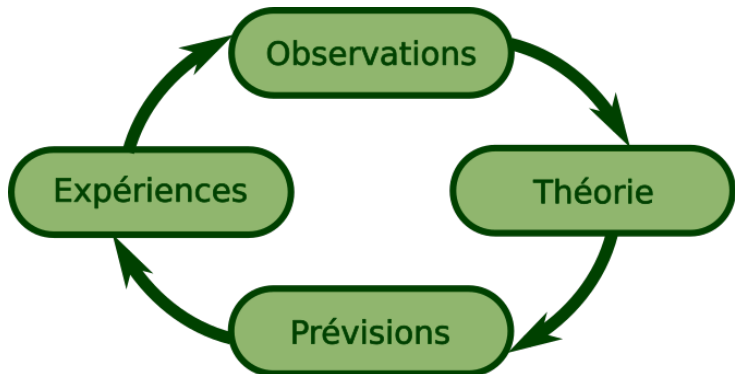
Schéma



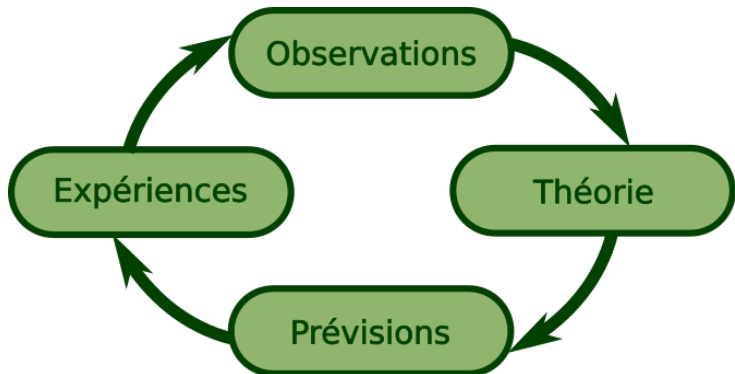
Schéma



Schéma



Schéma



Méthode « récente » échelle de l'humanité (Galilée, Descartes), mais qui a permis des progrès fulgurants !

Pourquoi ça marche

- Observations et mesures objectivées : indépendantes des opinions
- Falsifiable¹, donc vérifiable
- Universelle

1. Un énoncé est falsifiable « si la logique autorise l'existence d'un énoncé ou d'une série d'énoncés d'observation qui lui sont contradictoires, c'est-à-dire, qui la falsifieraient s'ils se révélaient vrais »

Pourquoi ça marche

- Observations et mesures objectivées : indépendantes des opinions
 - Falsifiable¹, donc vérifiable
 - Universelle
- ⇒ Approche rationnelle et regard sceptique sur le monde
Permet d'utiliser la science comme outil pour prendre des décisions
« politiques » basées sur des connaissances.

1. Un énoncé est falsifiable « si la logique autorise l'existence d'un énoncé ou d'une série d'énoncés d'observation qui lui sont contradictoires, c'est-à-dire, qui la falsifieraient s'ils se révélaient vrais »

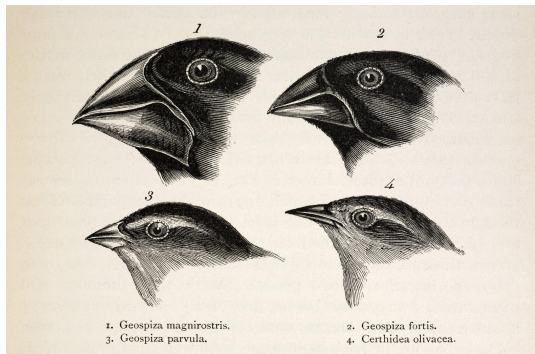
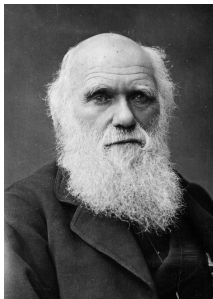
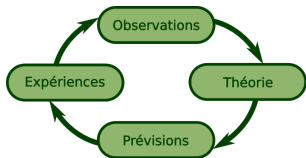
Pourquoi ça marche

- Observations et mesures objectivées : indépendantes des opinions
 - Falsifiable¹, donc vérifiable
 - Universelle
- ⇒ Approche rationnelle et regard sceptique sur le monde
Permet d'utiliser la science comme outil pour prendre des décisions « politiques » basées sur des connaissances.

Même si aujourd'hui crise inquiétante de confiance dans la science (et c'est une raison de plus d'être irréprochable dans vos pratiques !)

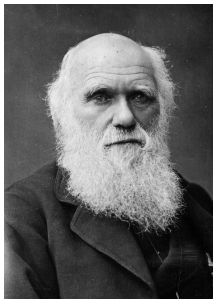
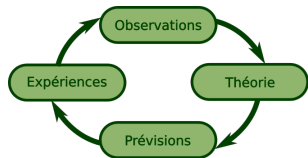
1. Un énoncé est falsifiable « si la logique autorise l'existence d'un énoncé ou d'une série d'énoncés d'observation qui lui sont contradictoires, c'est-à-dire, qui la falsifieraient s'ils se révélaient vrais »

Exemple de découverte scientifique : l'évolution



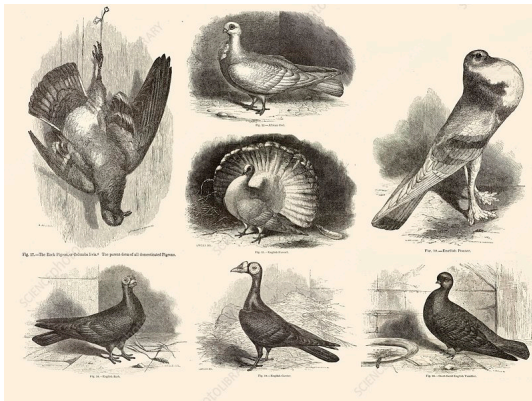
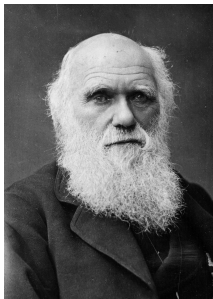
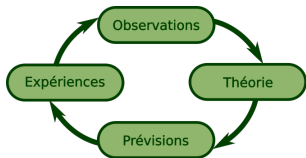
1831-1836 : Voyage à bord du *Beagle* et observations

Exemple de découverte scientifique : l'évolution



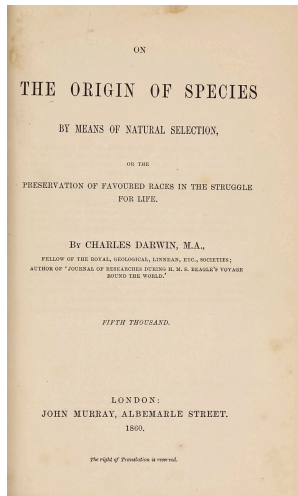
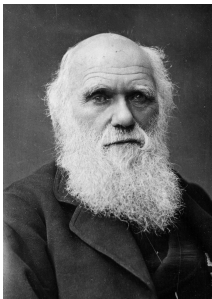
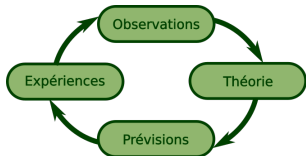
Aout 1836 : « such facts would undermine the stability of Species. » – début d'une théorie

Exemple de découverte scientifique : l'évolution



1855 : Début d'expériences sur les pigeons

Exemple de découverte scientifique : l'évolution

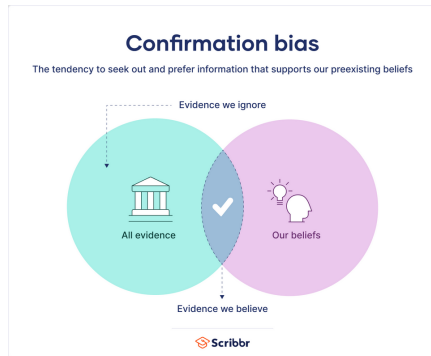


1859 : Publication de son livre *L'Origine des Espèces*

Mais comment être sûr ?

La méthode scientifique n'est pas infaillible !

- Biais de confirmation très fort
- Beaucoup d'autres obstacles (biais du survivant, biais d'échantillonnage...)
- Parfois mauvaise foi



- Pour valider un savoir, on recourt à la validation pour les pairs
- Pour être validé, un savoir doit être **reproductible**.

Et en entreprise ?

- Vous tomberez (souvent) sur du code legacy, dont plus personne ne sait que faire, et qui est irremplaçable ⇒ potentiellement grave à long terme ! Risque de sécurité, de fiabilité, ...
 - Il faut convaincre les clients que votre logiciel fait ce que vous dites qu'il fait
 - Collaborer avec des collègues implique d'avoir confiance en leur code... et leur donner confiance dans le vôtre !
- ⇒ Votre code doit aussi être **reproductible** !

Et en entreprise ?

- Vous tomberez (souvent) sur du code legacy, dont plus personne ne sait que faire, et qui est irremplaçable ⇒ potentiellement grave à long terme ! Risque de sécurité, de fiabilité, ...
 - Il faut convaincre les clients que votre logiciel fait ce que vous dites qu'il fait
 - Collaborer avec des collègues implique d'avoir confiance en leur code... et leur donner confiance dans le vôtre !
- ⇒ Votre code doit aussi être **reproductible** !

Qu'est-ce que ça veut dire ?

Reproductibilité, répétabilité et répliquabilité

Définitions – Association for Computing Machinery, 2016

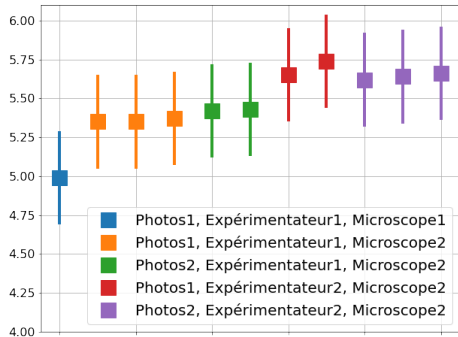
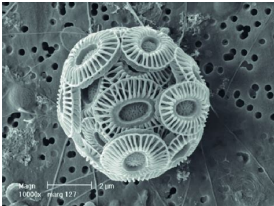
Repeatability (*Same team, same experimental setup*) : The measurement can be obtained with stated precision by the same team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same location on multiple trials.

Replicability (*Different team, same experimental setup*) : The measurement can be obtained with stated precision by a different team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same or a different location on multiple trials.

Reproducibility (*Different team, different experimental setup*) : The measurement can be obtained with stated precision by a different team, a different measuring system, in a different location on multiple trials.

Exercice pratique de définitions

On veut mesurer le diamètre de coccosphères. Pour vérifier la fiabilité de la mesure, on répète la mesure avec deux microscopes différents, 2 personnes différentes qui prennent les photos, et 2 personnes différentes qui prennent les mesures sur ces photos.



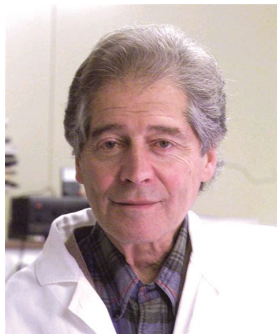
Que peut-on conclure ? Peut-on comparer différentes conditions de croissance ?

La mémoire de l'eau – un échec célèbre de reproductibilité

Une expérience répétable, répétée, non reproductible.

La mémoire de l'eau – un échec célèbre de reproductibilité

Une expérience répétable, répétée, non reproductible.



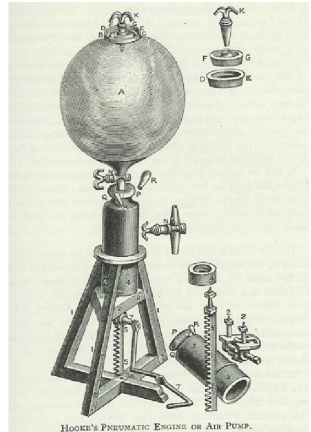
- Expérience publiée dans *Nature* (1988) (et fait la Une du *Monde*).
- **Répétée** devant un panel d'experts missionnés par la revue.
- Échec de **réplication** des résultats par d'autres laboratoires.
- D'autres expériences (mieux designées) ne permettent pas de retrouver les conclusions : pas de **reproductibilité**.

Autres échecs historiques

- Expérience de Stanford en 1971 (toujours considéré comme fondatrice en psychologie)
- Traitement du SIDA en 1985 (battage médiatique)
- Bargh et al. (2001) (on High-Performance-Goal Priming Effects), échec de réplication en 2013 (Harris et al. 2013)
- Reinhart et Rogoff 2011 : erreur « bête » sur excel détectée en 2013, mais grosse influence politique
- Hydroxychloroquine, ivermectine contre le covid en 2020
- ...

Demi réussite : de l'importance de la documentation

- Pompe à vide par Boyle en 1660
- Pas d'internet : Huygens voyage à Londres pour réussir à reproduire la pompe (et **répète** le résultat)
- **Réplique** l'expérience, mais note des bulles à air non suffisamment documentées précédemment
- Réalise une expérience différente pour montrer que l'hypothèse explicative n'est pas la bonne : résultat **non-reproductible**
- 2015 : reproduction des 2 expériences et explications 'définitives'



Solution to the long-standing puzzle of Huygens' "anomalous suspension", Nauenberg 2015

En résumé, la méthode n'est pas infaillible !

- Tout ce qui est publié n'est pas forcément vrai :
 - problème de la *p-value*, et du *p-hacking*
 - parfois mauvaise méthode
 - il existe des fraudes
- La pression à la publication empire ce problème :
 - Pression pour publier sans être totalement sûr
 - On ne vérifie pas les publications des autres, car ça ne permet pas de publier soi-même

⇒ Crise de reproductibilité

Qu'est-ce que la p -value?

Qu'est-ce que la p -value ?

Dans un test statistique, la p -value est la probabilité pour un modèle statistique donné sous l'hypothèse nulle d'obtenir une valeur au moins aussi extrême que celle observée. (*Wikipédia*)

- Souvent 5% \implies Donc 5% de faux positifs
- Si on recommence la même expérience suffisamment de fois, on trouve toujours un cas « positif » (*p-hacking*)
- Il existe des corrections à appliquer (ex : correction de Bonferroni), mais ce n'est pas toujours respecté

Crise de reproductibilité

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P.A. Ioannidis

Cette publication elle-même est remise en question sur sa reproductibilité !

- Problème exacerbé par un certain sensationnalisme dans la presse : une « connaissance » devient répandue avant d'avoir été vérifiée (HCQ, vaccins et autisme, ...)
- Les retractations existent mais arrivent souvent trop tard

Chiffres

Sondage (! pas forcément représentatif) sur environ 1 500 chercheurs par *Nature*, en 2016 (<https://www.nature.com/articles/533452a>) :

- 90% pensent qu'il y a une crise de reproductibilité
- 70% ont déjà échoué à reproduire le résultat d'un pair
- 50% ont déjà échoué à reproduire leur propre résultat

Chiffres

Sondage (! pas forcément représentatif) sur environ 1 500 chercheurs par *Nature*, en 2016 (<https://www.nature.com/articles/533452a>) :

- 90% pensent qu'il y a une crise de reproductibilité
- 70% ont déjà échoué à reproduire le résultat d'un pair
- 50% ont déjà échoué à reproduire leur propre résultat

→ Pourquoi ?

- Manque de documentation
- Manque de communication des données
- Mauvais design d'expériences
- Malchance ou *p-hacking* ?

La seule définition à retenir

Un travail non proprement documenté
est un travail à jeter

Conseils pratiques

Avant de coder

- Choisir un langage de programmation
 - Adapté au projet
 - Maintenu
 - Documenté, avec une communauté
 - Qui propose éventuellement des paquets d'intérêt
- Choisissez l'organisation / l'architecture de votre projet – et documentez-la
- Mettez en place un environnement adapté pour ne pas avoir de versions concurrentes ou inconnues des paquets que vous utilisez (conda, venv)

Bonnes pratiques d'écriture de code

- Utiliser des noms de variables qui ont du sens
- Respecter les usages de formatage
- Commenter quand nécessaire
- Ne pas dupliquer du code (utiliser des fonctions/classes dès que c'est approprié)
- **Documenter les usages prévus et comportements attendus** (pas directement dans le code)
- Écrire des tests unitaires (et les utiliser)
- Ajouter un log // prévoyez les données de sorties

"Be lazy, write good code"

→ Choisissez un IDE et apprenez à le connaître, ça vous fera gagner beaucoup de temps !

Bonnes pratiques – Commenter son code ?

```
function biggestNum(arr) {  
  let num = Number.NEGATIVE_INFINITY  
  for (let i = 0; i < arr.length; i++) {  
    if (arr[i] > num) {  
      num = arr[i]  
    }  
  }  
  return num  
}
```

Bonnes pratiques – Commenter son code ?

```
// Gets biggest numbers in a list
function biggestNum(arr) {
  // Initialize the candidate number to a smallest value
  let num = Number.NEGATIVE_INFINITY

  for (let i = 0; i < arr.length; i++) {
    // Check if the current number is greater than the candidate
    if (arr[i] > num) {
      num = arr[i]
    }
  }

  // Return the candidate number.
  return num
}
```

Bonnes pratiques – Commenter son code ?

```
function getBiggestNumber(numbers) {  
  let biggestSoFar = Number.NEGATIVE_INFINITY  
  
  for (let i = 0; i < numbers.length; i++) {  
    if (numbers[i] > biggestSoFar) {  
      biggestSoFar = numbers[i]  
    }  
  }  
  
  return biggestSoFar  
}
```

Bonnes pratiques – Quelles métadonnées ?

Si vous générez des données, par simulation/modélisation, analyses de données pré-existantes ou autres, vous **devez** documenter ces données pour qu'elles soient compréhensibles et utilisables !

- Auteurs, datation
- Version du code utilisée
- Labels et format de vos données de sortie
- Informations sur les données d'entrée (seeds ?)
- Toute autre information utile sur la méthode utilisée, les hypothèses utilisées et celles testées...

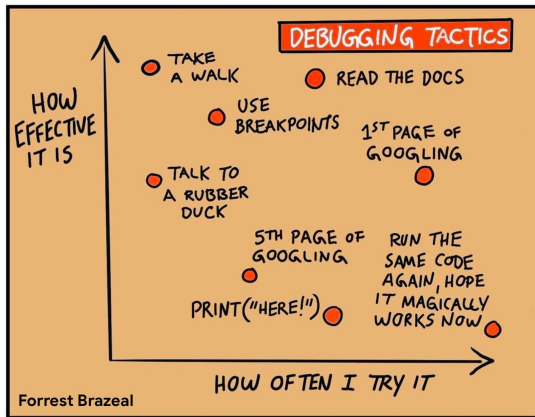
Bonnes pratiques – collaboration autour du code

Travailler avec quelqu'un ajoute du chaos, et donc des erreurs. Mais c'est aussi la possibilité de se relire mutuellement pour corriger !

- Utiliser Git proprement :
 - Une branche par feature
 - Commits unitaires et descriptifs
 - Utiliser les outils de revue de code
- Penser aux relecteur · rices lorsqu'on écrit son code
- Tester le code de ses collaborateur · rices, et leur fournir les outils pour qu'ils testent le vôtre
- Se coordonner en groupe (méthode AGILE)

Quelques techniques de debug

Debugging Tactics



Quand on utilise du code

- Logger le commit utilisé, la date
- Documenter les données d'entrées utilisées
- Documenter tout le protocole suivi
- Éventuellement stocker les données de sortie
- Décrire et analyser les résultats : que montre votre « expérience », qu'est-ce que le résultat vous permet de conclure et pourquoi

→ Il existe des cahiers de laboratoire en ligne

! Vérifiez que si vous éteignez tout et relancez de 0 vous arrivez au même résultat ! (*surtout avant une démonstration à un public/ un oral noté*)

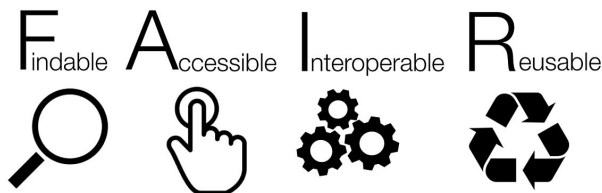
Exemple du machine learning

- Documenter ce que sont les jeux d'entraînements, de test, et de validation
- Enregistrer les hyper-paramètres
- Si l'entraînement est long, enregistrer le modèle entraîné
- Éviter les biais en équilibrant les données d'entraînement ou en adaptant la méthode au déséquilibre
- Sauvegarder aussi les seeds utilisées !
- Dans les cas plus extrêmes et si on veut comparer des temps d'exécution : noter les spécifications de la machine utilisée !

Principe FAIR

Un principe général à suivre pour assurer une meilleure reproductibilité

- **Findable** : Données enrichies de métadonnées, assorties d'un identifiant unique – et des indications claires pour trouver les données
- **Accessible** : Données et métadonnées stockées en open-access, accessible par les machines et les humains
- **Interoperable** : Format utilisable par tous et documenté, respect des normes et conventions
- **Reusable** : Fournir des documents additionnels pour expliquer la génération et l'utilisation des données, et associer une licence libre si possible



Conclusion

- La reproductibilité de votre travail est **essentielle**
- Repousser à plus tard une mise au propre, c'est souvent ne pas la faire
- Ne pas s'assurer de la reproductibilité, c'est un risque important pour votre projet et votre vie professionnelle future
- Prenez soin de la méthode avec laquelle vous acquérez vos connaissances, gardez votre biais de confirmation à l'œil, et vérifiez vos sources !
- Slides disponibles à juliette.luisel.li/data/reproductibility.pdf

Bon courage pour vos projets !