



# Recherche reproductible : état des lieux

Le nouveau Track « Recherche Reproductible » du Master MVA

Laurent Oudre

9 mars 2023

# Le projet SaclAI School

- Projet de massification de l'enseignement en Intelligence Artificielle porté par l'Université Paris Saclay (AMI CMA) débuté à la rentrée 2022
- Nombreuses initiatives visant à élargir les formations en IA et sciences des données : tutorat (TutorIA), bourses d'excellence (MixtIA), plateformes d'acculturation (BrevetIA), de recrutement (RecrutIA), de reconversion (ReconvertIA)....



## Le master MVA



- Master « Mathématiques, Vision et Apprentissage » piloté par l'ENS Paris Saclay
- Plus de 200 étudiants par an, issus d'universités et des grandes écoles
- Formation aux métiers de la recherche pour les organisations publiques ou privées en France et à l'international
- Domaines variés: l'industrie, le marketing digital, la défense etc...
- Existence d'un Track Santé dédié aux étudiants souhaitant utiliser les outils de l'IA dans la santé numérique

## Le nouveau Track « Recherche Reproductible »

- Création à la rentrée 2022 d'un nouveau Track autour des thématiques de la science reproductible
- Un constat : ces thématiques sont fondamentales dans le cadre d'un parcours de formation en machine learning
- L'historique du Centre Borelli : la revue IPOL



- Revue scientifique en traitement des images créée en 2010 par Jean-Michel Morel
- Articles de recherche associés à du code source (reviewé) et d'un démonstrateur en ligne permettant de relancer les expériences mais également de tester sur ses propres données

# Le nouveau Track « Recherche Reproductible »

- Deux nouveaux modules :
  - Semestre 1 : Fondamentaux de la recherche reproductible et du logiciel libre
  - Semestre 2 : Projet de recherche reproductible
- Un nouveau séminaire “Panorama Recherche Reproductible” ouvert à tous les étudiants du MVA
- Track RR : valider les deux modules (NB : les étudiants doivent en valider 8 en tout) + assister aux séminaires

# Module 1 : Fondamentaux de la RR et du logiciel libre

- Notions vues en cours:
  - Logiciel libre, logiciel open source et licences
  - Qu'est-ce que la reproductibilité ? Les différentes définitions et leurs implications
  - Aspects légaux
  - Comment promouvoir la reproductibilité dans vos travaux de recherche ?
- Travaux pratiques
  - Etude d'un package logiciel existant (modèle économique, licences, audience...)
  - Analyse de la reproductibilité du package logiciel (depend-il d'autres librairies ? Le code a-t-il été reviewé ? Causes et conséquences des non-reproductibilités)
  - Etude d'un article de recherche et étude de sa reproductibilité (code source disponible ? Reviewé ? Correspondance article/code ?)

[http://mcolom.perso.math.cnrs.fr/pages/courses/MVA\\_FOSS/index.html](http://mcolom.perso.math.cnrs.fr/pages/courses/MVA_FOSS/index.html)

## Module 2 : Projet de Recherche Reproductible

- Principe:
  - Prendre un article de recherche existant, avec code source
  - Le comprendre, auditer le code fourni et son adéquation avec l'article
  - Ecrire un nouveau rapport scientifique complet sur la méthode et son implementation et construire un démonstrateur
- Délivrable :
  - Soumission à la revue IPOL

[http://mcolom.perso.math.cnrs.fr/pages/courses/MVA\\_RRP/index.html](http://mcolom.perso.math.cnrs.fr/pages/courses/MVA_RRP/index.html)

## Le bilan (provisoire)

- Une quinzaine de participants pour le module 1, répartis en 4 groupes
  - Sujets divers : reconnaissance de parole, restauration d'images, algorithmes d'optimisation, modèles génératifs profonds
  - Retours très positif des étudiants
- Une dizaine de participants pour le module 2 (toujours en cours)
  - A priori 4 publications (article+code+demo) parfaitement reproductibles