

Projet de recherche

Résumé

L'année 2023 aura été marquée par une exposition médiatique forte des outils numériques dérivés des recherches en intelligence artificielle. Au-delà des mythes, il est important, en tant que scientifique, d'évaluer la pertinence des utilisations proposées. Mon projet de recherche, dans le contexte de ma candidature au PhD Track a pour objectif 1) **de mieux comprendre** les principes théoriques et algorithmiques à la base du fonctionnement et de l'utilisation des IA, 2) **de découvrir les utilisations en cours**, à la fois dans le monde des entreprises biotechnologiques et celui de la recherche académique et 3) **de définir les contours d'un projet de thèse interdisciplinaire** qui allierait les potentialités nouvelles issues des algorithmes d'apprentissage machine (IA) avec des problématiques de modélisation des systèmes biologiques. Celles-ci seraient mises en application dans un contexte de recherche en génomique fonctionnelle des champignons, thématique de recherche du laboratoire d'accueil prévu pour la thèse.

Partie 1 : Mieux comprendre les principes de fonctionnement des IA

ChatGPT, DeepL, MidJourney, etc. Ces outils dérivés des méthodes d'apprentissage automatique ont été considérablement exposés dans les médias ces derniers mois. Si certaines personnes voient dans leurs utilisations des potentialités immenses, d'autres en sont très méfiantes. **En tant que scientifique, mon souhait est de mieux comprendre les principes théoriques et algorithmiques, qui sont à la base du fonctionnement de ces outils.** Comment se passe les procédures d'entraînement des algorithmes ? Quelle est l'impact de la qualité et la quantité des données utilisées ? Quelles sont les ressources informatiques nécessaires à leur bon fonctionnement ? Des utilisations sont-elles possibles dans un contexte de recherche en biologie ? Je propose de collecter des éléments de réponses à ces questions au cours de ma première année de Master (année universitaire 2023 – 2024). Pour cela, j'ai trouvé un stage de deux mois dans un laboratoire de biologie expérimentale qui commence à utiliser des stratégies d'analyses de données biologiques dérivées du "deep learning". L'objectif est d'étudier *sans a priori*, les signaux contenus dans des cartes de fréquences de contacts mesurées dans des génomes bactériens (technique expérimentale Hi-C). **Il s'agit d'un stage de Bioinformatique qui me permettra de réaliser une première mise en application concrète des méthodes d'apprentissage automatique.** Cette première expérience me permettra d'identifier d'éventuels points de blocage pour mon travail avec les IA, par exemple théoriques (nécessité de renforcer mes connaissances en mathématiques, statistiques ou algorithmiques) ou techniques (nécessité de renforcer mes compétences en programmation ou d'apprendre à utiliser des infrastructures informatiques particulières). Dans un second temps, je profiterai de mon stage de deuxième année de Magistère (2 mois également) pour mener une enquête dans plusieurs entreprises. **Il s'agira pour moi de rencontrer des ingénieurs et des chercheurs en charge des services "data science" et "IA" d'entreprises privées et publiques.** Je les questionnerais concernant leurs usages des méthodes d'apprentissage automatique, afin de mieux comprendre les plus-values associées pour leurs activités et également les limitations rencontrées. **Je m'intéresserai également aux évolutions des métiers qui émergent de ces usages.** Ce travail d'enquête sera réalisé avec l'accompagnement et le soutien financier de ma future mentore (Gaëlle Lelandais), pour la prise en charge entre autres de mes frais de déplacement. Nous identifierons également ensemble les personnalités à rencontrer et effectuerons toutes les démarches administratives pour l'organisation logistique des entretiens. Les résultats de cette enquête seront mis en forme et valorisés par la rédaction d'un rapport (éventuellement une publication scientifique).

Partie 2 : Définir les contours d'un projet de thèse interdisciplinaire

Je souhaite que mon projet de thèse s'inscrive dans la thématique "développement soutenable" de ce PhD Track. **En cohérence avec l'expertise de recherche de mon laboratoire d'accueil, je souhaite m'intéresser aux champignons.** Ce sont en effet des organismes clés de nos écosystèmes, de par leur capacité à recycler la matière organique et leurs nombreuses interactions avec les plantes. Ils sont essentiels à la survie de la forêt et au maintien de l'équilibre de tous les biotopes terrestres. Également, les champignons trouvent leur importance aussi bien dans notre alimentation avec la production de vin, de pain ou encore la conservation de la viande en saucisson et du lait en fromage, que dans la santé et l'industrie grâce à leur production de nombreuses molécules d'intérêts thérapeutiques ou industriels comme les antibiotiques par exemple, ou des enzymes utilisés dans les lessives pour enlever les tâches de gras. (*Les champignons redécouverts*, Silar et Malagnac ; *Les champignons pourront-ils sauver le monde ?* Rizzo et Sipp, reportage Arte). De plus, les biotechnologies développées autour des champignons permettent de contribuer à 10 des 17 objectifs du développement soutenable des Nations Unies (Figure 1). A titre plus personnel, je suis particulièrement intéressée par les activités d'entreprises de biotechnologies qui mettent au point des procédés permettant d'utiliser les champignons afin de produire des analogues du cuir pour l'industrie du textile ou encore des analogues de composites synthétiques du bois (MycoWorks). D'autres encore proposent une alternative à la viande de poulet (Quorn). Aussi, les champignons sont utiles pour dépolluer les sols grâce à la mycoremédiation et leur capacité à dégrader des éléments chimiques et polluant (Meyer et al. 2015). Mon objectif serait donc **de réaliser des modélisations *in silico* des systèmes fongiques**, soit à l'échelle macroscopique (interactions entre des ensembles d'espèces de champignons et leur environnement), soit à l'échelle génétique/génomique (interactions entre les composants cellulaires, au cours des différentes étapes de développement des structures différenciées (type hyphes ou fructifications). Ces modélisations reposeraient sur **l'entraînement d'algorithmes d'apprentissage non supervisé**, à partir de sources de données hétérogènes. Faire un bilan des données disponibles et choisir les algorithmes et les méthodes de modélisation pertinentes à implémenter sera l'objectif principal de ma deuxième année de Master (année universitaire 2024 – 2025). Je prévois dans ce contexte de **solliciter un stage de 6 mois dans une équipe de modélisation des systèmes biologiques** (au savoir-faire principal associé aux mathématiques / statistiques), pour une première expérience pratique sur le sujet.

Partie 3 : La thèse

Il y a quelques mois encore, l'idée de réaliser une thèse n'était pas une évidence pour moi. Curieuse de découvrir le monde du travail, je n'imaginais pas continuer mes études aussi longtemps. C'est une discussion avec ma future mentore qui m'a fait changer d'avis. Elle m'a expliqué sa vision de la thèse : une première expérience professionnelle réelle, au cours de laquelle je pourrais développer mon propre projet, en cohérence, bien sûr, avec les intérêts de mon laboratoire d'accueil, mais aussi avec une très grande autonomie. **Réussir à obtenir par moi-même, cette bourse du PhD Track, serait une réelle opportunité de pouvoir débiter mon travail au laboratoire en toute indépendance.** Je sais que je pourrais compter sur les conseils et le soutien de ma directrice de thèse, et que je pourrais bénéficier de l'expertise de laboratoire pour évaluer la pertinence des questions biologiques auxquelles je tenterais de répondre. Toutefois, le projet aura l'immense intérêt de m'être très "personnel", car il émergera de réflexions menées depuis ma première année de Master. **Je me sentirais responsable de sa réussite et en ce sens, l'obtention de cette bourse m'engage et me motive tout particulièrement.**