

Partager



Email destinataire :

Vous devez renseigner un email valide

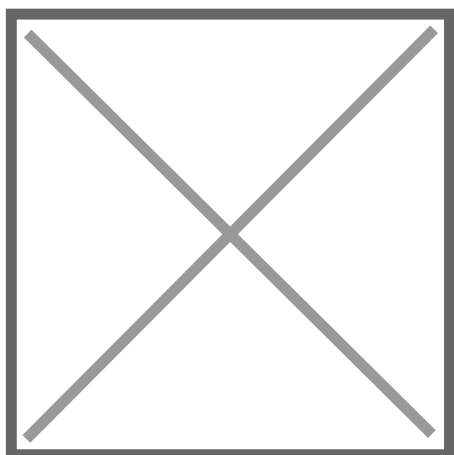
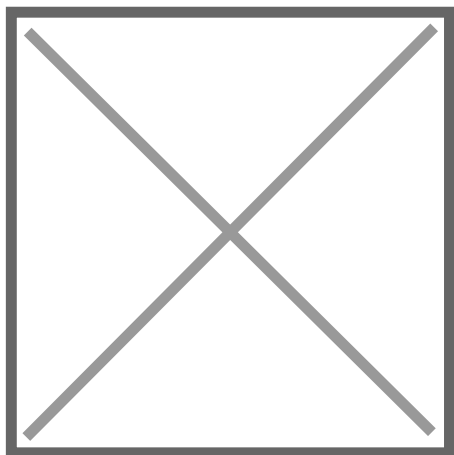
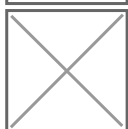
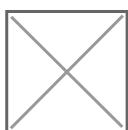
Envoyer

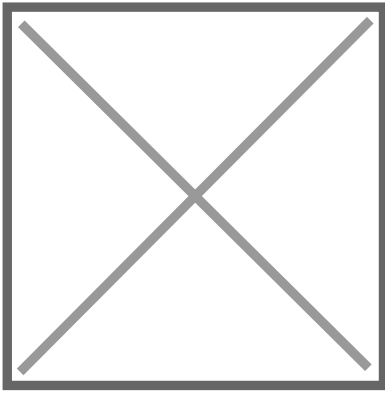
Une erreur est survenue. Merci de réessayer ultérieurement

Le mail de partage a bien été envoyé.

Ajouter au comparateur

M2 Bioinformatique, Biostatistiques et applications de l'intelligence artificielle (BIBS-IA)





Master
Mention Bio-informatique
Langue(s) d'enseignement
Français

Le **M2 BIBS-IA du Master Bioinformatique** forme des **experts en recherche et développement** à l'intersection de la biologie, de l'informatique et des mathématiques. Les étudiants acquièrent des compétences recherchées en **intelligence artificielle**, gestion et analyse de données et modélisation dans divers domaines de la bioinformatique - génomique, biologie structurale, biologie évolutive, santé numérique. Ces connaissances sont appliquées en sciences du vivant, médecine, environnement et agronomie.

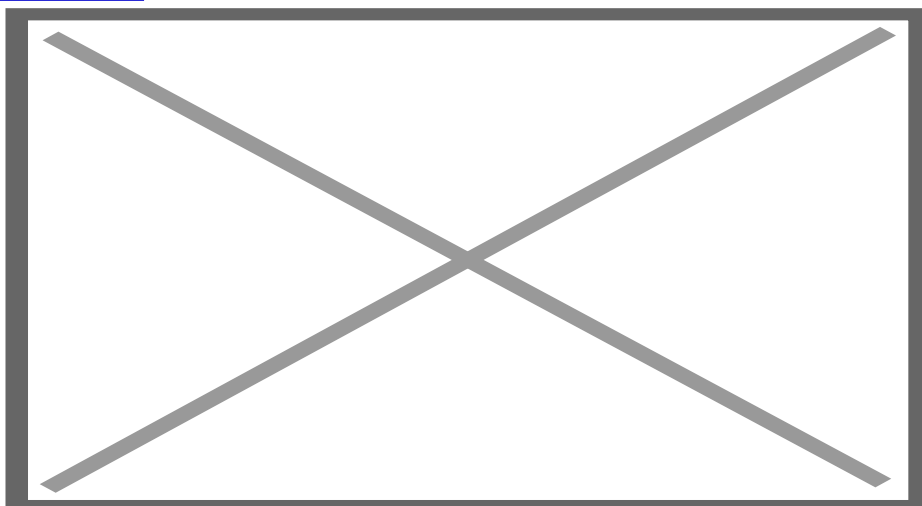
Le M2 BIBS-IA est construit dans la continuité du M1 BIBS-IA. La section "Profil" indique les pré-requis attendus.

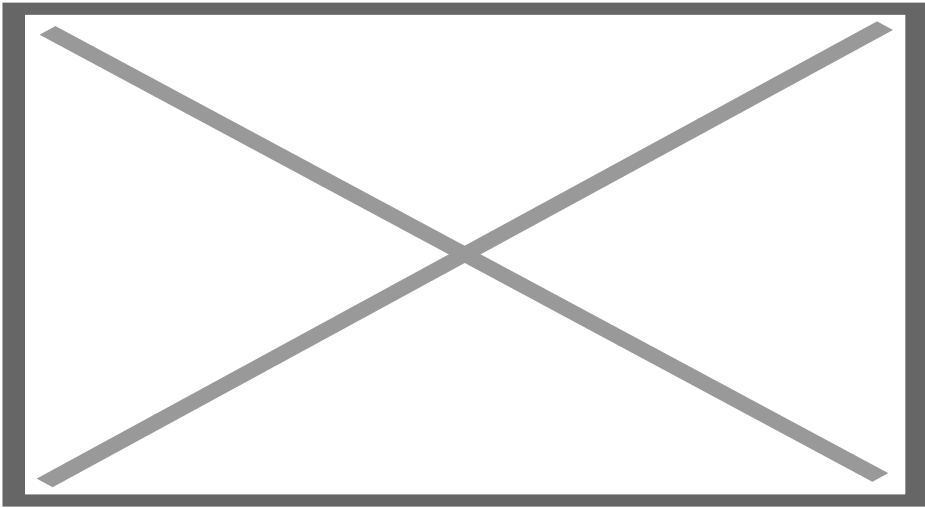
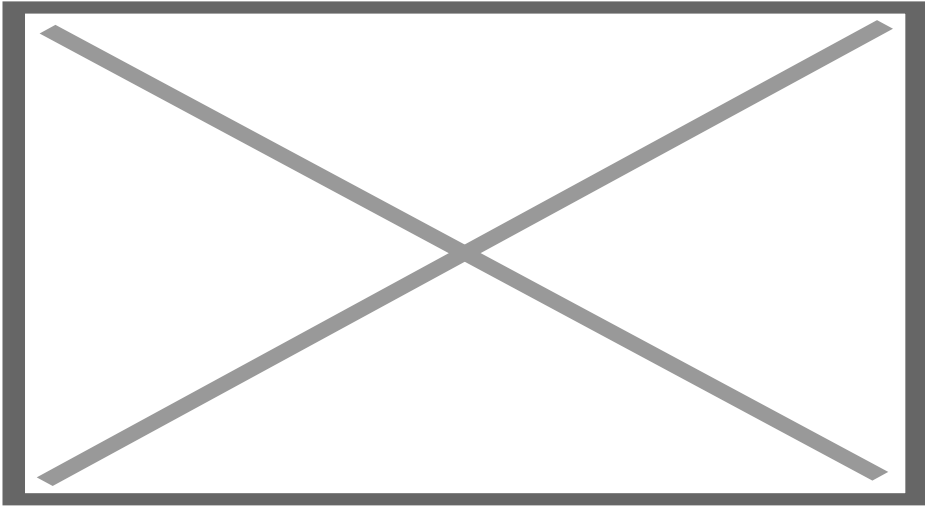
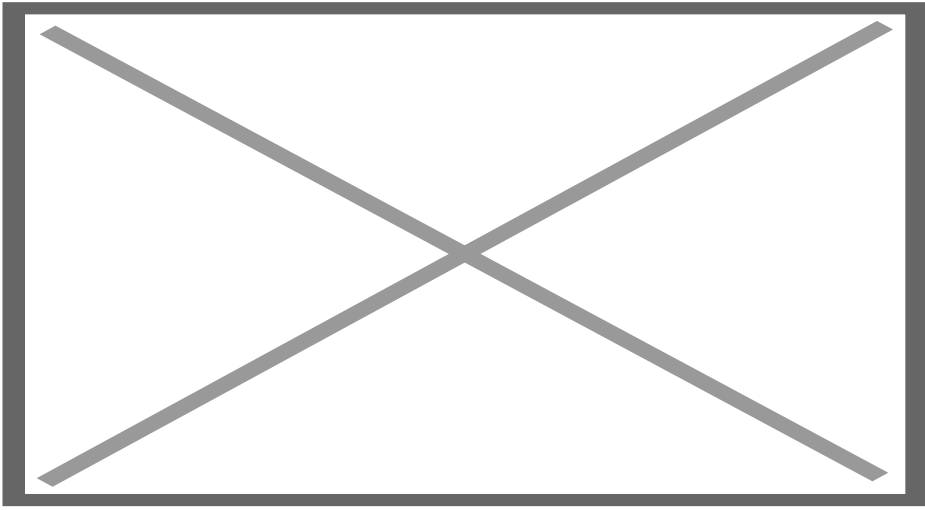
Dans le parcours BIBS-IA, l'**intelligence artificielle et la science des données** sont très présentes et traitées à la fois d'un point de vue **fondamental** en mathématiques et informatique, et **appliqué** en lien fort avec les spécificités des données biologiques et biomédicales.

Des **projets et hackathons pluridisciplinaires** sont dispensés par des experts du milieu académique et des entreprises. Les étudiants accèdent à un vaste panel de modules d'enseignement au choix, leur permettant de se spécialiser (applications de l'Intelligence artificielle en biologie structurale, en génomique, en contexte clinique, en agronomie ...).

Un stage obligatoire de 4 à 6 mois assure une future bonne insertion des étudiants dans le monde du travail, en recherche ou en industrie.

[Télécharger la fiche détaillée](#)
[Je candidate](#)







■ ■

Informations

- [Présentation](#)
- [Admission](#)
- [Programme](#)

Présentation

Compétences

Les compétences acquises en fin de M2 BIBS-IA sont reprises de la fiche nationale du Master de Bioinformatique ([RNCP 38964](#)).

- **Comprendre les enjeux actuels de la complexité et de la diversité en biologie**, et sciences du vivant en général et de contribuer activement à la recherche et à l'innovation dans les secteurs académique, pharmaceutique, biotechnologique, agro-industriel et de la santé ;
- **Relever les défis résultant de l'évolution et du développement des technologies de production de données à haut débit**, issus de sources multiples et hétérogènes ;
- **Maîtriser les approches computationnelles et l'intelligence artificielle** pour analyser, modéliser et interpréter des données biomédicales à grande échelle.
- **Piloter des projets bioinformatiques** de manière autonome, en environnement multidisciplinaire, en développant des solutions innovantes avec des outils et langages informatiques adaptés.
- **Expérimenter, développer et tester de nouvelles méthodes**, pour explorer des hypothèses en sciences du vivant à la frontière entre recherche fondamentale et appliquée.

Objectifs pédagogiques de la formation

Le M2 BIBS-IA est axé sur les **besoins en recherche et développement de haut niveau des entreprises et des organismes de recherche en bioinformatique, et biostatistiques**.

Cette formation pluridisciplinaire répond à cette demande en formant des spécialistes à l'interface des disciplines : biologie, informatique et mathématiques.

A l'issue du M2 BIBS-IA, les étudiants **maîtrisent plusieurs langages de programmation**, un grand nombre de **méthodes d'intelligence artificielle, d'algorithmique, de statistiques et de sciences des données**, pour stocker, modéliser et analyser des **masses de données** génomiques et biomédicales. Ils approfondissent par ailleurs des domaines de la bioinformatique (biologie des systèmes, biologie structurale, génomique fonctionnelle) et appliquent les méthodes apprises sur des projets variés, portant sur les sciences du vivant et du médical, ainsi que les sciences de l'environnement et l'agronomie.

Les enseignements du M2 BIBS-IA comportent 6 grands blocs de connaissances et compétences.

- **Bloc 1 - Gestion de grands volumes de données et optimisation.** Ce bloc comprend trois unités d'enseignement relatives à l'optimisation des traitements de données diverses (bases de données relationnelles et données graphe). Un projet portant sur des données réelles massives permet une mise en application des connaissances acquises.
- **Bloc 2 - Génomique haut débit.** Ce bloc comprend deux unités d'enseignement avec un focus en génomique comparée et un autre en génomique des populations. Un accent important porte sur l'application des méthodes statistiques dans ces domaines.
- **Bloc 3 - Analyse de données et IA pour la biologie.** Il comprend trois unités d'enseignement depuis l'acquisition de connaissances fondamentales en IA comme les modèles de *machine learning* jusqu'à des applications de l'IA en biologie intégrative et en analyse d'images.
- **Bloc 4 - Intégrité scientifique et reproductibilité.** Il comprend deux unités d'enseignement de préprofessionnalisation sous la forme d'un projet de reprohackathon et d'une unité d'aide à l'insertion professionnelle. Les étudiants acquièrent dans ce bloc l'ensemble des bonnes pratiques attendues et relatives à la science ouverte, l'intégrité scientifique, l'éthique et la sobriété énergétique (impact écologique des pratiques du numérique et notamment de l'Intelligence Artificielle).
- **Bloc 5 - Applications Avancées de l'Intelligence Artificielle.** Les étudiants font un choix de 4 unités dans le bloc 5. Ce bloc comprend six unités d'enseignement propres au parcours M2 BIBS-IA et dédiées à diverses applications de l'IA. Elles incluent par exemple un "data camp", hackathon de *machine learning* sur données réelles (enseignement commun avec l'école polytechnique et Télécom Paris) ou encore, de la fouille de données en agronomie (commun avec AgroParisTech). A ces unités s'ajoutent des unités libres, notamment issues d'autres parcours de Master (Informatique, biologie-santé...).

Débouchés

Professionnels

Après Master + Doctorat : chercheur ou enseignant-chercheur

Ingénieur Bioinformaticien (Recherche et Développement)

Expert en gestion de données

data scientist

Expert science des données

Poursuite d'études

Doctorat en Bioinformatique

Chercheur/chercheuse en R&D ou expert·e en modélisation et analyse de données dans des entreprises ou laboratoires de pointe. Data Scientist, Data Analyst, Ingénieur·e en Machine Learning dans des secteurs innovants (tech, finance, santé, énergie, etc.) ;

Agenda

Début de la formation (date indicative)

03/10/2025

Fin de la formation (date indicative)

04/09/2026

Admission

Capacité d'accueil

Places

24

Public visé et prérequis

Les enseignements du M2 BIBS-IA sont en majorité dispensés en français, la langue française doit donc être maîtrisée.

Le M2 BIBS-IA est construit dans la continuité du M1 BIBS-IA. Il regroupe trois profils d'étudiants :

- les **étudiants ayant validé le M1 BIBS-IA** qui sont tous admis en M2 BIBS-IA
- les **étudiants d'AgroParisTech suivant la dominante IODAA** ([*De l'information à la décision par l'analyse et l'apprentissage*](#)) qui font tous leur année en double diplôme avec le M2 BIBS-IA
- (selon les places disponibles) les étudiants ayant effectué un M1 équivalent au M1 BIBS-IA c'est-à-dire ayant acquis des connaissances et compétences équivalentes à la fois en biologie, mathématiques et informatique (voir le programme du M1 BIBS-IA pour plus d'information).

Les étudiants ayant déjà un diplôme de Master de Bioinformatique ne sont pas éligibles, les étudiants ayant déjà un diplôme de Master dans un domaine autre que la bioinformatique (biologie ou informatique ou mathématiques) sont éligibles à une entrée en M1. Leurs candidatures seront néanmoins considérées comme moins prioritaires que celles des étudiants sortants de licence.

Période(s) de candidature

Plateforme Inception

Pièces justificatives

Obligatoires

Lettre de motivation.

Tous les relevés de notes des années/semestres validés depuis le BAC à la date de la candidature.

Curriculum Vitae.

Facultatives

Attestation de français (obligatoire pour les non francophones).

Attestation de niveau d'anglais (obligatoire pour les non anglophones) ou Résultats test GMAT/GRE.

Descriptif détaillé et volume horaire des enseignements suivis depuis le début du cursus universitaire.

Dossier VAPP (obligatoire pour toutes les personnes demandant une validation des acquis pour accéder à la formation) <https://www.universite-paris-saclay.fr/formation/formation-continue/validation-des-acquis-de-lexperience>.

Fiche de choix de M2 (obligatoire pour les candidats inscrits en M1 à l'Université Paris-Saclay) à télécharger sur <https://urlz.fr/i3Lo>.

Document justificatif des candidats exilés ayant un statut de réfugié, protection subsidiaire ou protection temporaire en France ou à l'étranger (facultatif mais recommandé, un seul document à fournir) :

- Carte de séjour mention réfugié du pays du premier asile
- OU récépissé mention réfugié du pays du premier asile
- OU document du Haut Commissariat des Nations unies pour les réfugiés reconnaissant le statut de réfugié
- OU récépissé mention réfugié délivré en France
- OU carte de séjour avec mention réfugié délivré en France
- OU document faisant état du statut de bénéficiaire de la protection subsidiaire en France ou à l'étranger.

Programme

Bloc 1 - Gestion de grands volumes de données et optimisation

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
Projet données massives	2	Semestre 1	2		18						

Projet données massives

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 2

Travaux pratiques : 18

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Savoir développer en Python, se servir de git

Programme / plan / contenus

- Architecture client/serveur
- Prise en main du framework Django
- Focus sur certaines fonctionnalités de Django
- Projet

Objectifs d'apprentissage

Concevoir une application web permettant à plusieurs profil d'utilisateurs de consulter et enrichir les données.

Avoir conscience de enjeux en terme de sécurité informatique.

Concevoir une base de donnée répondant aux besoins de l'application, et la volumétrie des données.

Savoir comprendre et se mettre à la place des utilisateurs, et anticipier les besoins.

Conduire et mener à terme un projet de développement informatique.

Apprendre à travailler en équipe.

Organisation générale et modalités pédagogiques

Une série de cours/TD initiaux permet aux étudiants de se familiariser avec le framework Django, l'architecture client/serveur et un framework css. Au cours de cette UE, les étudiants réalisent un projet en groupe de 3-4 où ils doivent concevoir et développer une application web interrogeant et enrichissant une base de données. Le suivi du projet se fait sous forme d'entretiens hebdomadaires avec un enseignant.

Bibliographie

<https://bbrancot.pages.pasteur.fr/prog-web-client-server/>
<https://bbrancot.pages.pasteur.fr/django-course/>

Nature de l'évaluation

Évaluation Continue et Finale

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
<u>Bases de Données Avancées</u>	2	Semestre 1	10	10							

Bases de Données Avancées

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 10

Travaux dirigés : 10

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Savoir implémenter en SQL la création de schémas, tables, clés primaires et étrangères, l'insertion/suppression/mise à jour de données et l'interrogation à l'aide des opérateurs Select-From-Where-Group by-Having.
Avoir des notions en Algèbre relationnelle (opérateurs, requêtes)

Programme / plan / contenus

- Stratégies de placement des données sur le disque (stockage des données)
- Index : types et structures
- Niveaux d'optimisation d'une BD : schéma, requête
- Plan de requêtes, algorithmes existants pour l'implémentation des opérateurs
- Réécriture de requêtes

Objectifs d'apprentissage

L'objectif de cette UE est de présenter les techniques les plus avancées pour optimiser l'accès aux données d'une base de données : Savoir optimiser une requête SQL ; savoir optimiser un schéma de bases de données.

Organisation générale et modalités pédagogiques

Séances hebdomadaires de Cours-TD

Bibliographie

Database Management Systems
de Raghu Ramakrishnan, Johannes Gehrke
McGraw Hill, 3ème édition

Database Tuning. Principles, Experiments, and Troubleshooting
de Dennis Shasha, Philippe Bonnet
Morgan Kaufmann

Nature de l'évaluation

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
Graphes biologiques et optimisation	2	Semestre 1	12	6							
■											

Graphes biologiques et optimisation

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 12

Travaux dirigés : 6

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Programme d'algorithmique du M1 BIBS

Programme / plan / contenus

- Théorie de graphes. : bases, coloration, graphes planaires. - Algorithmique des graphes : parcours, plus courts chemins, arbres coouvrant de poids minimum, flots. Illustrations sur des problèmes en bioinformatique. - Théorie de la complexité : problèmes P, NP, NP-complets; exemples dans les graphes. - Optimisation et Approximabilité.

Objectifs d'apprentissage

Acquérir les bases et les compétences en théorie et algorithmique des graphes et en complexité

Organisation générale et modalités pédagogiques

Séances hebdomadaires de Cours-TD

Bibliographie

Claude Berge. Graphes. Dunod.

Nature de l'évaluation

Evaluation Continue non Intégrale

Bloc 2 - Génomique haut débit

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
Génomique des populations	2	Semestre 1	10		10						

Génomique des populations

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 10

Travaux pratiques : 10

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Bases en analyse de séquences, notions de statistiques et probabilités, et familiarité avec Python.

Programme / plan / contenus

Cette UE propose une **introduction aux méthodes utilisés en génétique des populations à travers une application biologique concrète**. Après avoir présenté les concepts clés (mutation, recombinaison, généalogie des individus) les étudiantes et étudiants explorent un cas de métissage entre populations. Ce fil conducteur permet de découvrir différentes **approches d'inférence**, basées sur des simulations (Approximate Bayesian Computation, Machine Learning,...) et sur des **modèles probabilistes** (Hidden Markov Models). Les étudiants apprennent à manipuler un simulateur de séquences versatile et des bibliothèques Python pour analyser, visualiser et interpréter des données, souvent représentées sous forme de graphes de généalogie.

Objectifs d'apprentissage

Acquérir des compétences en **analyse et inférence à partir de données complexes de type graphe**, dans le contexte de la **génétique des populations**. Les étudiants utilisent des outils informatiques et des modèles probabilistes pour explorer et interpréter des scénarios issus de simulations.

Organisation générale et modalités pédagogiques

L'UE alterne cours magistraux et travaux pratiques. Les cours présentent les concepts fondamentaux de la génétique des populations et les approches d'inférence. Les TP permettent aux étudiantes et étudiants de manipuler des données simulées et d'analyser des cas réels, en appliquant les méthodes étudiées pour explorer différents scénarios et interpréter les résultats.

Bibliographie

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
<u>Génomique comparée</u>	2	Semestre 1	10		10						

Génomique comparée

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 10

Travaux pratiques : 10

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Connaître un langage de script

Avoir des connaissances de base en génomique

Programme / plan / contenus

Concepts d'orthologie, core et pan génomes.

Détermination du contenu en gènes d'un génome

Application à la comparaison de génomes complets

Analyse des génomes : Duplications, Pertes et Gains de Gènes, Transferts horizontaux

Génomique comparée (Diagramme de Venn / Conservation de la synténie)

Objectifs d'apprentissage

Savoir manipuler, analyser et interpréter les données génomiques pour répondre à une question biologique

Cette UE vise à donner aux étudiants une vision synthétique de la constitution et de l'évolution des génomes procaryotes et eucaryotes.

Organisation générale et modalités pédagogiques

L'enseignement se fait en salle informatique et consiste à analyser des jeux de données extraits de la bibliographie ou produits dans les laboratoire du campus afin d'appréhender en situation réelle tous les aspects de l'analyse des données génomiques et de la comparaison des génomes.

Nature de l'évaluation

Evaluation Continue non Intégrale

Bloc 3 - Analyse de données et IA pour la biologie

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
Analyse d'Images en Biologie	2	Semestre 1	8		12						

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
■											

Analyse d'Images en Biologie

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 8

Travaux pratiques : 12

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Aucun

Programme / plan / contenus

Décrire et expliquer les bases théoriques de l'image numérique et de son interprétation :

1. Byte, bit, pixel, formats d'image, rang dynamique, histogramme, LUTs, opérations mathématiques sur les images.
2. Échantillonnage
3. Espace de Fourier, opérations dans l'espace des fréquences.
4. Segmentation, types de bruit et filtres (espace réel et Fourier)
5. Projection et rétroprojection
6. Visualisation 3D, utilisation d'ImageJ (Fiji), utilisation de plug-in, programmation Macro
7. Convolution et corrélation

Extraire des données quantitatives de l'image :

1. Segmentation par seuillage et introduction aux approches d'apprentissage profond pour la segmentation
2. Densitométrie
3. Granulométrie

Traiter et analyser les signaux :

1. Signaux périodiques (cristaux 2D et/ou structures empilées - hélices). Calcul de paramètres de signaux périodiques
2. Rapport signal sur bruit, contrastes, entropie
3. Débruitage

Réaliser des reconstructions 3D à partir de projections :

1. Algorithmes de reconstruction en imagerie : théorème de la section central, rétroprojection dans l'espace de Fourier, rétroprojection pondérée, techniques de reconstruction algébriques.
2. Effets du sous-échantillonnage et du manque d'information.

Automatiser des tâches par programmation Macro

Objectif de la formation

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
<u>Analyse de données omiques et biologie intégrative</u>	2.5	Semestre 1	15	5	20						

Analyse de données omiques et biologie intégrative

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 15

Travaux pratiques : 20

Travaux dirigés : 5

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Bonne maîtrise du logiciel R; Principes des techniques d'analyses de données multivariées (ACP, MDS, etc.); Statistique inférentielle (estimation, tests d'hypothèses, etc.); Techniques de séquençage haut débit (RNAseq, ChIPseq, etc.)

Programme / plan / contenus

Différents types de données issues des approches expérimentales haut débit (omiques) seront traitées (par exemple RNAseq, ChIPseq, spectrométrie de masse). Une première partie de l'enseignement sera consacrée à l'analyse exploratoire des données multivariées, puis seront abordées différentes problématiques telles que l'analyse différentielle (RNAseq), la détection de pics (peak calling, ChIPseq) et l'inférence de réseaux de gènes.

Objectifs d'apprentissage

- Modéliser les données omics à l'aide de variables aléatoires adaptées et poser un cadre statistique inférentiel nécessaire à l'analyse de ces données
- Implémenter des méthodes d'estimations de paramètres dans les modèles
- Mettre en oeuvre les méthodes proposées sur des jeux de données réelles

Organisation générale et modalités pédagogiques

Les séances (3 h 00) se déroulent en salle informatique, avec une alternance de phases de cours et de phases de mises en applications pratiques. Des évaluations des acquis de compétences sont proposées régulièrement aux étudiants sous la forme de questionnaires sur Internet (auto-évaluations non notées). Le travail en groupe est

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
Modèles pour le Machine Learning	2.5	Semestre 1	15		15						

Modèles pour le Machine Learning

ECTS : 2.5

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 15

Travaux pratiques : 15

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Estimation dans le modèle linéaire, Estimation par maximum de vraisemblance, Programmation statistique avec R

Programme / plan / contenus

Modèles de régression et classification supervisée, paramétriques et non-paramétriques;
Ré-échantillonnage bootstrap, validation croisée et évaluation de performances (estimation de l'erreur de généralisation, courbes ROC);
Sélection de variables, choix de modèles;
Arbres de décision (CART, Forêts aléatoires);
Méthodes d'apprentissage en grande dimension (LASSO, SVM)

Objectifs d'apprentissage

Génération, évaluation et comparaison de modèles de machine learning:

Être capable d'utiliser des méthodes variées d'apprentissage statistique et de les interpréter sur des données, d'implémenter les méthodes étudiées avec R;
Savoir mettre en place une stratégie de choix de modèle(s); Savoir évaluer et comparer les performances des modèles d'apprentissage.

Organisation générale et modalités pédagogiques

Chaque séance inclut une étude de cas permettant la mise en oeuvre des méthodes enseignées et l'analyse statistique des données associées

Bibliographie

An Introduction to Statistical Learning, with applications in R, Gareth, Witten, Hastie and Tibshirani

Nature de l'évaluation

Évaluation Continue non Intégrale

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
Reprohackathon	3	Semestre 1	11		19						



Reprohackathon

ECTS : 3

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 11

Travaux pratiques : 19

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Savoir développer en python.

Nature de l'évaluation

Evaluation Continue non Intégrale

[Mise en situation
professionnelle](#)

2 Semestre 1 10 15

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
■											

Mise en situation professionnelle

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 10

Travaux dirigés : 15

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Aucun

Programme / plan / contenus

L'enseignement combine :

- Des ateliers pratiques de rédaction de CV et de préparation aux entretiens ;
- Une introduction aux bonnes pratiques de recherche éthiques et à la responsabilité environnementale ;
- Une junior conférence organisée par les étudiants sur deux jours, leur permettant de se mettre en situation réelle de communication scientifique et de gestion d'événement académique.

Objectifs d'apprentissage

Cette UE vise à préparer les étudiants à leur insertion professionnelle et à les sensibiliser aux enjeux éthiques et environnementaux de la recherche en bioinformatique.

À l'issue de l'UE, les étudiants sauront :

- Identifier les principaux débouchés de la bioinformatique dans la recherche publique et privée ;
- Préparer leur orientation entre la poursuite en doctorat et une carrière d'ingénieur ;
- Concevoir un CV et se préparer à un entretien professionnel ;
- Adopter de bonnes pratiques de recherche et une réflexion éthique et responsable sur leur activité scientifique.

Organisation générale et modalités pédagogiques

L'UE alterne cours et exercices centrés sur la mise en pratique.

L'évaluation repose sur la participation active, les productions personnelles (CV, simulations d'entretien) et l'implication dans l'organisation de la junior conférence.

Nature de l'évaluation

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
Méthodologie en Recherche Clinique et Statistiques	2	Semestre 1	10	10							



Méthodologie en Recherche Clinique et Statistiques

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 10

Travaux dirigés : 10

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Introduction à la recherche clinique et épidémiologique
(UE de M1 BIBS)

Objectifs d'apprentissage

Comprendre le développement d'un médicament /
dispositif médical dans son environnement technique et
réglementaire

Connaitre les acteurs et les métiers de la recherche
clinique

Voir comment un parcours BIBS-IA peut intéresser les
entreprises des sciences de la vie

Comprendre les grandes fonctions statistiques liées à la
recherche clinique.

Organisation générale et modalités pédagogiques

Cours magistral et ateliers interactifs.

Nature de l'évaluation

Evaluation Continue Intégrale

[LLM pour les
séquences
protéiques](#)

2

Semestre 1

10

10

LLM pour les séquences protéiques

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 10

Travaux pratiques : 10

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Maîtriser un langage de programmation. Bases de statistiques et de bioinformatique structurale.

Programme / plan / contenus

Introduction de la révolution de l'IA dans la prédiction des structures 3D des protéines, impact sur l'annotation grande échelle des métagénomomes. Introduction aux LLM de protéines. Analyse manuelle de quelques prédictions structurales et classification des prédictions en fonction de leur score de confiance. Analyse à l'échelle d'un protéome. Utilisation des embeddings de LLMs (ESMfold et ProtTrans) et de méthodes de bioinformatique pour l'analyse de séquences et structures pour interpréter et classifier les cas prédits avec mauvais score de confiance (ie, utiliser d'autres approches pour donner de la valeur aux cas prédits avec faible confiance, typiquement écartés des approches classiques). Tirer partie de cette analyse pour discuter des limites des LLM pour explorer des territoires inconnus (souvent le cas avec les données de métagénomiques venant d'environnements extrêmes ou sous-échantillonnés). Cette analyse offrira un cas pratique idéal pour appliquer certaines méthodes statistiques vues en cours (lois de mélange, ...).

Objectifs d'apprentissage

- Comprendre les modèles LLM entraînés sur les séquences protéiques : architecture et entraînement
- Connaître les applications principales des modèles LLM des protéines (prédiction de structure, fonction, ...)
- Visualiser et interpréter les embeddings d'un LLM
- Analyse de données à grande échelle (analyse d'un protéome, puis de plusieurs dizaines de protéomes)
- Mobiliser ses compétences en statistiques et programmation pour répondre à une question concrète et

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
UE libre Catalogue S1	2.5	Semestre 1									
■											

UE libre Catalogue S1

ECTS : 2.5

Semestre calendaire : Semestre 1

Langue d'enseignement

Français

Enseignement à distance

non

Nature de l'évaluation

Evaluation Terminale

[Advanced
Machine
Learning](#)

2

Semestre 1

10

10

Advanced Machine Learning

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 10

Travaux dirigés : 10

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

Avoir suivi un cours d'apprentissage automatique couvrant en particulier les apprentissages supervisé et non supervisé.

Programme / plan / contenus

- L'induction classique avec données i.i.d. Problème. Quel type de garantie peut-on vouloir et obtenir ? Conditions pour les obtenir.
- Quand on change la distribution pour obtenir de meilleurs résultats : boosting, apprentissage actif, ré-échantillonnage, Learning Using Privileged Information (LUPI)
- Apprentissage semi-supervisé. Apprentissage avec des agents communiquant entre eux : co-training, blending, analogie, effet tunnel cognitif, distillation
- Quand la distribution P_X change entre apprentissage et test. Adaptation de domaine et co-variate shift.
- Quand la distribution conditionnelle $P_{Y|X}$ change. Tracking, apprentissage multi-tâches ; apprentissage par transfert, zero-shot learning, modèles de fondations, Invariant risk minimization
- Apprentissage continu avec curriculum. Comment changer la tâche, comment mesurer la difficulté d'un nouvel exemple, comment choisir une bonne séquence d'exemples ou de tâches d'apprentissage.
- Apprentissage en-ligne : approches théorique et pratique. Effets de séquence. Oubli catastrophique.

Objectifs d'apprentissage

L'objectif du cours est d'étudier l'apprentissage dit « hors distribution », c'est-à-dire quand les données d'apprentissage ne sont pas issues de la même distribution, ni même éventuellement du même domaine, que les données test. C'est par exemple le cas de l'apprentissage avec des données issues d'une certaine région (e.g. l'Europe), ou à propos d'un certain organisme (e.g. un certain type de bactérie), et dont le résultat, l'hypothèse apprise, sera utilisée pour une autre région

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
Data Camp	2	Semestre 1	20		10						

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
■											

Data Camp

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 20

Travaux pratiques : 10

Langue d'enseignement

Anglais

Enseignement à distance

non

Prérequis

Une bonne connaissance de python et de l'écosystème de calcul scientifique en python est requise (numpy, scipy, scikit-learn). Une connaissance des bases de l'apprentissage automatique et des algorithmes de classification est aussi nécessaires.

Programme / plan / contenus

Jour 1 : Préparation et manipulation des données

- * Introduction du workflow (VSCode, Python, git, GitHub, tests, ...)
- * Cours avancé sur Pandas
- * Exercices GitHub : NumPy et Pandas

Jour 2 : Pipelines de ML et évaluation de modèles

- * Scikit-learn avancé : ColumnTransformer et pipelines
- * Traitement parallèle avec joblib
- * Généralisation et validation croisée
- * Exercices sur scikit-learn
- * Introduction à CodaLab et présentation des challenges

Jour 3 : Métriques et gestion des données déséquilibrées

- * Présentation des différentes métriques de ML
- * Problèmes liés aux métriques avec des données déséquilibrées
- * Approches d'apprentissage automatique pour traiter les jeux de données déséquilibrés
- * Travail sur les challenges de données

Jour 4 : Données complexes

- * Feature engineering et encodage avancé des variables catégorielles
- * Travailler sur des séries temporelles et du signal
- * Inspection des modèles : Partial Dependence Plots, Feature Importance
- * Travailler sur un data challenge

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
Fouille de données pour le vivant	2	Semestre 1	10		10						

Fouille de données pour le vivant

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 10

Travaux pratiques : 10

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Des bases en algorithmique et programmation (idéalement en Python) sont nécessaires.

Programme / plan / contenus

Sont abordées dans ce module des éléments théoriques permettant de comprendre les questions fondamentales de l'apprentissage non supervisé. Y seront présentés et détaillés quelques modèles, ainsi que des éléments de méthodologie et de mise en pratique.

Objectifs d'apprentissage

La fouille de données (data mining) est un domaine de l'intelligence artificielle qui s'intéresse à l'émergence de connaissances nouvelles à partir de données. On parle également d'apprentissage automatique non supervisé (unsupervised learning). L'objectif de ce module est d'étudier les problématiques de l'apprentissage automatique non supervisé et les différentes catégories de méthodes qui ont été proposées pour y répondre avec des applications dans le domaine des sciences du vivant et l'environnement. On se place dans un cadre "non supervisé" lorsqu'on n'a pas d'exemples pour lesquelles on connaît ce que l'on cherche à prédire. Par exemple, on dispose d'une collection d'images. Certaines représentent des plantes et d'autres des animaux mais on ne sait pas à quelle "catégorie" appartient chacune de nos images. Parfois, il est possible et facile d'obtenir ce type d'informations et un apprentissage automatique dit "supervisé" (supervised learning) sera envisageable. Dans le cas contraire, comme souvent en sciences du vivant et de l'environnement où les "étiquettes" peuvent être chères, difficiles voire impossible à obtenir, il faudra aborder le problème de manière "non supervisée" (unsupervised learning). C'est ce que nous verrons lors de ce module.

Organisation générale et modalités pédagogiques

Ce module se découpe naturellement en deux parties.

Matières	ECTS	Semestre	Cours	TD	TP	Cours- TD	Cours- TP	TD- TP	A distance	Projet	Tutorat
Analyse d'images et intelligence artificielle	2	Semestre 1	10	10							

Analyse d'images et intelligence artificielle

ECTS : 2

Semestre calendaire : Semestre 1

Détail du volume horaire

Cours magistraux : 10

Travaux dirigés : 10

Langue d'enseignement

Français

Enseignement à distance

non

Prérequis

Bases d'analyse d'images biologiques; bases d'apprentissage profond

Programme / plan / contenus

L'enseignement est constitué autour d'une activité pratique d'apprentissage profond supervisé sur des images biologiques qui met en jeu des concepts et des techniques pertinentes et d'utilisation courante, et permet aux apprenants d'effectuer une telle activité avec une compréhension profonde des enjeux, en donnant une base solide des disciplines concernées pour minimiser les "boîtes noires". L'enseignement prévoit donc une importante partie de "projet", où les apprenants travaillent en groupe avec la supervision ponctuelle de l'équipe enseignante. L'activité de projet inclut les tâches suivantes : 1) Nettoyage et préparation des données ; 2) Choix d'un modèle existant pertinent ; 3) "Refinement" du modèle pour la tâche de prédiction souhaitée ; 4) Évaluation de la performance du nouveau modèle.

Objectifs d'apprentissage

- Comprendre les modèles d'apprentissage profond d'utilisation courante en analyse d'images (CNN et en particulier UNET) et mesurer la qualité de leurs performances

- Connaître les applications principales des modèles d'apprentissage profond pour l'analyse d'images biologiques (*e.g.* segmentation, classification, ...)

- Utiliser des modèles d'apprentissage profond pour l'analyse d'images biologiques à travers un logiciel (cellpose, YOLO, Stardist...)

- Analyser des images biologiques avec des modèles d'apprentissage profond pour en tirer une information biologique

Matières	ECTS	Semestre	Cours	TD	TP	Cours-TD	Cours-TP	TD-TP	A distance	Projet	Tutorat
Stage	30	Semestre 2									

Stage

ECTS : 30

Semestre calendaire : Semestre 2

Détail du volume horaire

Projet tutoré 45

Langue d'enseignement

Français

Enseignement à distance

non

Nature de l'évaluation

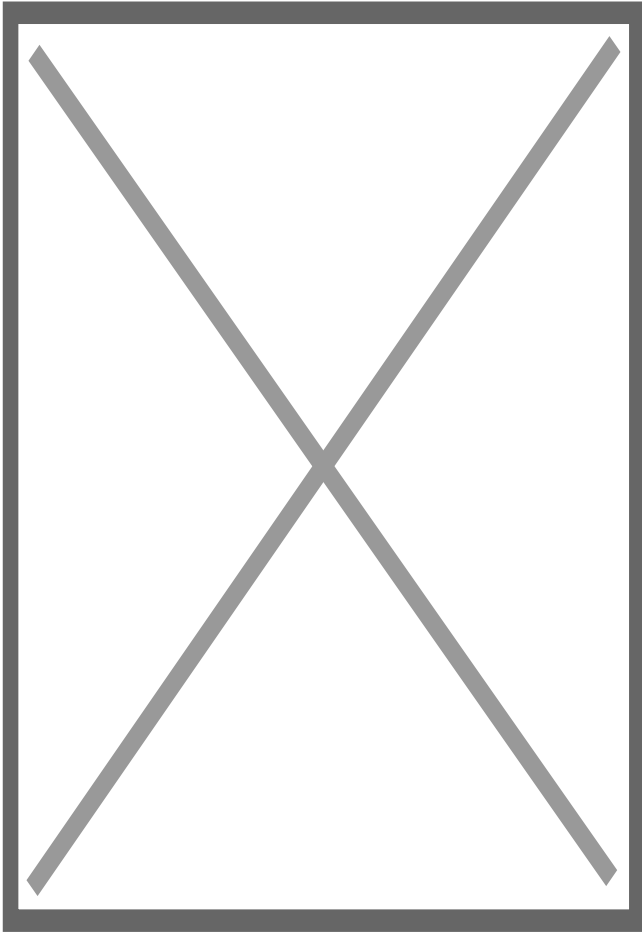
Evaluation Continue Intégrale

Lieu(x) d'enseignement

ORSAY

Campus de la formation

Javascript est requis pour visualiser cette carte.



Saclay Moulon

Saclay / Gif-sur-Yvette / Orsay

Bus 4606 / 4609 / 4611 / 5154

RER B Le guichet

[Voir](#)

Restaurant CROUS

Bibliothèque

Equipements sportifs

Résidence étudiante

Contact(s)

- **Fanny Pouyet**

MCF

<http://www.universite-paris-saclay.fr/formation/master/bio-informatique/m2-bioinformatique-biostatistiques-et-applications-de-lintelligence-artificielle-bibs-ia>