

Paris, le 13 juin 2025

Utilisation des IA génératives comme appui à la programmation et au scripting pour la biologie

Session 3 – Ateliers pratiques



Programme de la journée

Début	Fin	Titre	Intervenants	Lieu
08:30	09:00	Accueil des participants		
09:00	11:30	Session 1 – Introduction et enjeux de l'utilisation des IA génératives dans les pratiques bioinformatiques		Amphitheatre A1 - Hall aux Farines, campus des Grands Moulins, 5, rue Thomas Mann, 75013 Paris
09:00	09:05	Introduction à la journée	Bertrand Cosson, Jacques van Helden	
09:05	09:35	Outils IA pour le scripting et le codage : évolution et perspectives	David Janiszek	
09:35	10:05	Utilisation de grands modèles de langage génomique	Guillaume Gautreau	
10:05	10:35	Utilisation d'IA générative dans un processus pédagogique	Pierre Poulain	
10:35	11:00	Pause café		
11:00	11:30	Utilisation de Copilot pour le développement logiciel : principe, intérêt et limitations, cas d'étude	Nicolas Sabouret	
11:30	12:30	Session 2 – Retours d'expérience		
11:30	11:40	Adaptation pédagogique d'un pipeline RNAseq par Devin	Sandrine Caburet	
11:40	11:50	Présentation de Pléiade	David Janiszek	
11:50	12:00	Accélérer sans déraiper : maîtriser son code à l'ère de l'IA	Vincent Ranwez	
12:00	12:30	Table-ronde et interactions avec la salle		
12:30	14:00	Déjeuner (buffet)		Hall aux Farines
14:00	17:30	Session 3 – Ateliers pratiques		Salles d'informatique 273 + 281 +289 - Lamarck B, 35 rue Hélène Brion, 75013 Paris
14:00	15:30	Sessions pratiques par groupe	Imane Messak, Baptiste Rousseau, Thomas Denecker, Pierre Poulain, Gaëlle Lelandais, Vincent Ranwez, Fanny Casse, Jacques van Helden	
15:30	16:00	Pause café		
16:00	17:00	Sessions pratiques par groupe		
17:00	17:30	Conclusions de la journée et prochaines activités		Salle 1021, Bâtiment Sophie Germain, Pl. Aurélie Nemours, 75013 Paris

- Accès aux ressources numériques pour les ateliers pratiques
- Atelier 1 : appui sur une IA générative pour implémenter un script de bioanalyse
- Atelier 2 : appui sur une IA générative pour le développement logiciel
- Remerciements et financements

Accès aux ressources numériques pour les deux ateliers pratiques





Ressources bioinformatiques utilisées

Nom	URL
Ressource	URL
Programme et inscriptions	https://iabioscripting.sciencesconf.org/
Matériel de support (diapo, ateliers pratiques)	https://ifb-elixirfr.github.io/AI-for-scripting-bioanalysis/
Plateforme logicielle Pléiade utilisée pour l'atelier pratique	https://pleiade.mi.parisdescartes.fr/
Cluster IFB-core : interface OnDemand	https://ondemand.cluster.france-bioinformatique.fr/
Cluster IFB-core : demande de compte ou d'espace-projet <i>(pas nécessaire pour l'atelier, utile uniquement si vous désirez continuer à utiliser le cluster après la formation)</i>	https://my.cluster.france-bioinformatique.fr/

1. Dans un navigateur, ouvrez une connection vers l'interface OnDemand du cluster IFB :
<https://ondemand.cluster.france-bioinformatique.fr/>
2. Entrez le login et le mot de passe qui vous ont été fournis en début de séance.
Note : si vous disposez déjà d'un compte sur le cluster IFB-core, vous pouvez l'utiliser pour cet atelier pratique.
3. OnDemand vous présente un choix d'interfaces pour utiliser le cluster IFB
4. Choisissez l'interface qui vous convient selon votre langage préféré: RStudio (R) ou JupyterHub (Python, et éventuellement R).

OPEN
OnDemand
OnDemand provides an integrated, single access point for all of your HPC resources.

Pinned Apps A featured subset of all available apps

 JupyterLab: Core System Installed App	 RStudio Server: Core System Installed App	 RTrainer: Core System Installed App	 Desktop: Core System Installed App
---	---	---	--

Configuration des ressources de calcul

OnDemand vous permet de configurer les ressources de calcul qui vous seront attribuées pour votre session.

- Ajustez la durée à 4 heures, afin de pouvoir travailler pendant l'atelier.

Note : vous pourrez ultérieurement rouvrir une session JupyterLab ou RStudio et poursuivre votre travail (OnDemand rétablira l'état où se trouvait votre session quand elle s'est interrompue)

- Pour cet atelier pratique, laissez les autres paramètres inchangés. A l'avenir, vous serez éventuellement amenés à les adapter pour des analyses nécessitant des moyens de calcul spécifiques.
- Cliquez **Launch**.

JupyterLab: Core
This app will launch a JupyterLab server inside a SLURM job.

Reservation

Account

Partition

Number of CPUs

A least one CPU is required for the JupyterLab webserver

Amount of memory

Unit can be specified using the suffix M, G or T (default is megabytes).

GPUs

Number of GPUs

Number of hours

Launch

* The JupyterLab: Core session data for this session can be accessed under the data root directory.

RStudio Server: Core
This app will launch an RStudio server inside a SLURM job.

R version

Reservation

Account

Partition

Number of CPUs

You should reserve at least 2 CPUs to have enough resources for RStudio server and R.

Amount of memory

RStudio requires an amount of memory roughly equal to the amount of data you would like to load in your session. Unit can be specified using the suffix M, G or T (default is megabytes).

Number of GPUs

GPUs

Number of hours

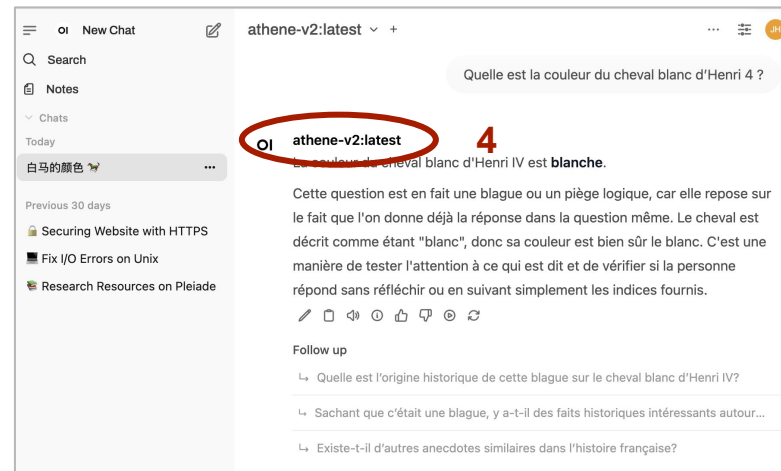
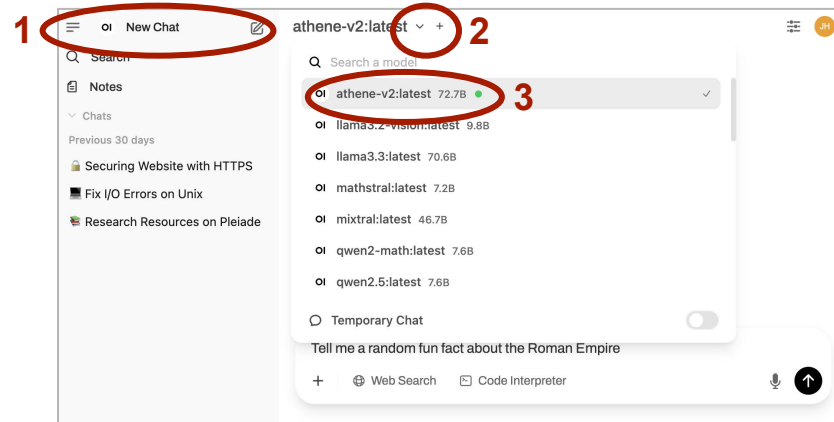
Launch

* The RStudio Server: Core session data for this session can be accessed under the data root directory.

Accès à la plateforme d'IA génératives Pléiade

- Connectez-vous à la plateforme Pléiade (login, mot de passe)
- <https://pleiade.mi.parisdescartes.fr/>
- Au cas où vous auriez déjà fait quelques essais préalables, cliquez sur “New chat” pour démarrer une nouvelle conversation.
- Pour cette session pratique, assurez-vous de choisir le modèle de langage
- Vous pouvez maintenant commencer à utiliser Pléiade.

Attention : Pléiade propose différents modèles de langage, mais pour ces travaux pratiques nous demandons à tous les participants d'utiliser le même modèle, pour éviter une surcharge du serveur (il doit recharger le modèle chaque fois qu'on en change). Vous aurez tout le loisir de tester les autres modèles de langage après l'atelier.



Atelier 1 : appui sur une IA générative pour implémenter un script d'analyse de données

Jeu de données – Profil transcriptomique oscillants durant le cycle cellulaire de la levure

Source des données:

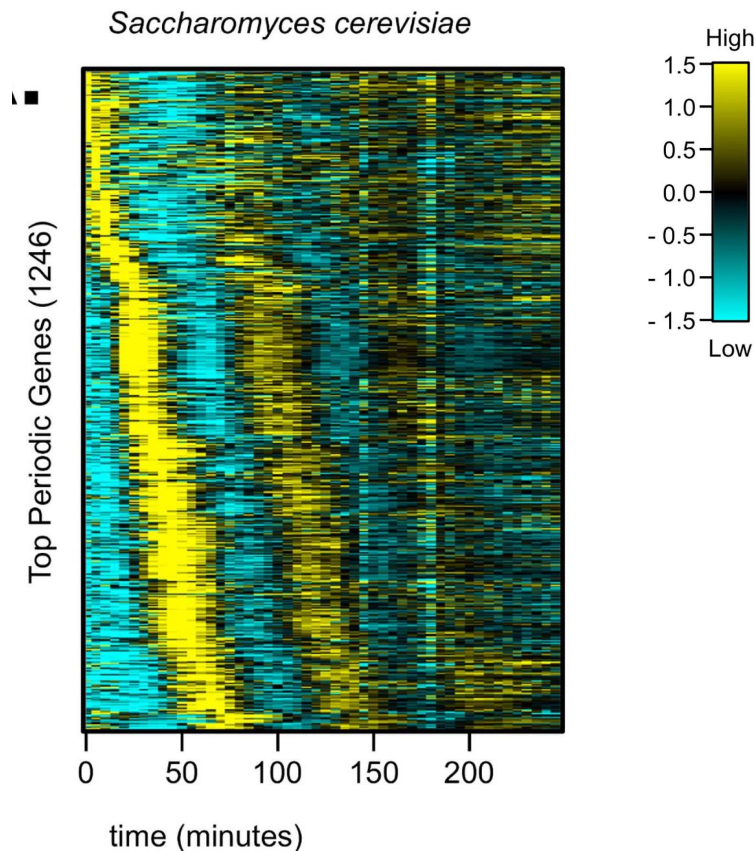
- Kelliher, C. M., Leman, A. R., Sierra, C. S. & Haase, S. B. Investigating Conservation of the Cell-Cycle-Regulated Transcriptional Program in the Fungal Pathogen, *Cryptococcus neoformans*. PLoS Genet 12, e1006453 (2016). DOI [10.1371/journal.pgen.1006453](https://doi.org/10.1371/journal.pgen.1006453)

Pré-traitement

- Effectué par Gaëlle Lelandais et les étudiants du Diplôme Universitaire Omique (DUO)
- Normalisation des comptages d'ARN (outil DeSEQ2)
- Les profils normalisés contiennent les $\log_2(\text{counts})$ de 1700 gènes présentant des oscillations d'expression durant le cycle cellulaire, sur 50 points temporels

But de l'exercice : utiliser une IA générative pour


- Charger un fichier de données tabulaire (profils d'expression normalisés)
- Calculer des statistiques marginales sur les lignes (gènes) et les colonnes (points temporels) du tableau.
- Produire une figure "heatmap" similaire à celle de l'article (ci-jointe).
- Exporter les statistiques marginales
- Exporter la figure en formats png et pdf.








- **Fournissez d'emblée le contexte à votre IA** : dites-lui votre but, donnez lui des précisions concernant le type de données, vos critères (documentation, structuration, ...)
- **Avancez progressivement** : plutôt que de demander d'emblée un script qui fait tout, commencez par demander un script qui fait la première étape (télécharger les données) puis la deuxième (charger les données), et avancez pas à pas.
- **Structurez le script** : demandez à votre IA de segmenter le code en implémentant des fonctions pour chacune des tâches différentes.
- **Effectuez des tests unitaires** : concevez un test de validation de chacune des fonctions et demandez à l'IA de vérifier que le code donne la bonne réponse.

Atelier 2 : appui sur l'IA générative pour le développement logiciel

Objectif : Utiliser l'intelligence artificielle pour améliorer vos scripts et analyses en bioinformatique.

 **Outils explorés :** Pleiade (différents modèles de langage), ChatGPT, Perplexity ou autres assistants IA

Ce que vous apprendrez :

-  Corriger automatiquement les erreurs de code
-  Optimiser le code pour plus d'efficacité et de lisibilité
-  Suggérer des améliorations selon les bonnes pratiques
-  Réussir les vérifications CI (intégration continue)
-  Générer fonctions, tests unitaires et documentation

L'IA comme alliée pour coder plus vite, mieux, et avec moins d'erreurs

ChatGPT

Avant de commencer les développements - forker projet

<https://github.com/IFB-ElixirFr/IA-BioSoftware-Atelier>

Pendant les développements

Écrire dans le fichier PROMPTS.txt pour ajouter les prompts que vous avez utilisés pour

À la fin de la session

Dans le README, ajouter

- l'outil utilisé (pléiade, chatGPT, copilot,...)
 - Le modèle utilisé
 - Le nombre de requêtes réalisées
 - Est-ce que le code passe les tests ?

Remerciements et financements

Remerciements

- David Janiszek pour la mise à disposition de la plateforme d'IA génératives **Pléiade**
- Le **cluster de calcul IFB-core**
- Le **Pôle informatique** UFR Sciences du Vivant de l'Université Paris-Cité
- **Reine Rigault** du Pôle Formation Professionnelle de l'Université Paris-Cité

Financements

Cette journée a été financée par

- La plateforme iPOP-up de l'Université Paris-Cité
- L'Institut Français de Bioinformatique (IFB)
- Le réseau-métier en bioinformatique MERIT
- Le Diplôme Universitaire "Création, analyse et valorisation de données omiques" (DUO)

L'Institut Français de Bioinformatique (IFB) a été fondé par les Programme d'Investissements d'Avenir subventionné par l'Agence Nationale de la Recherche ([RENABI-IFB, ANR-11-INBS-0013](#)) et par le programme France 2030 relatifs aux équipements structurants pour la recherche / EQUIPEX+ ([MUDIS4LS, ANR-21-ESRE-0048](#)).