



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*



Ressources, outils et formations pour l'IA

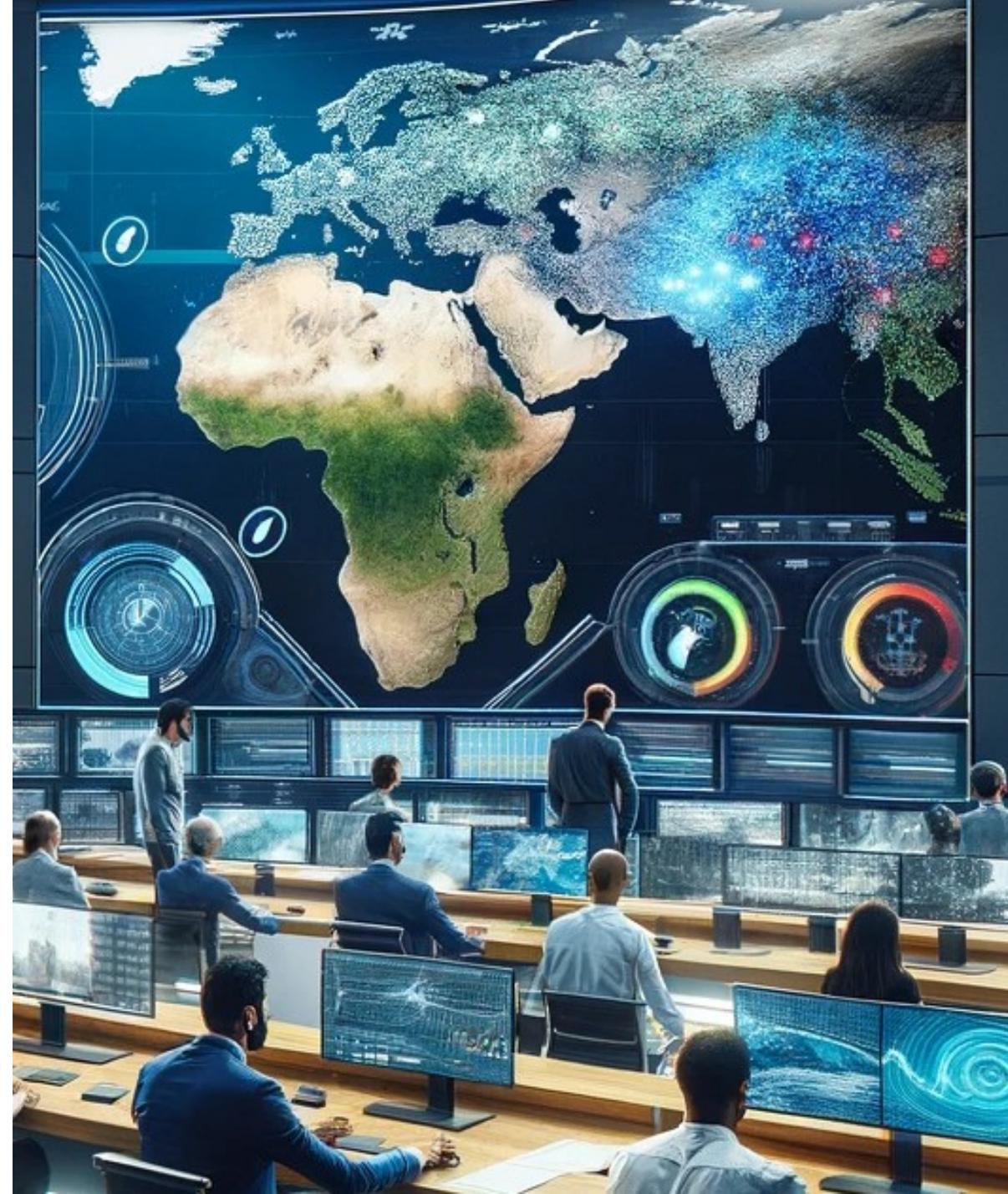
Léa Berthomier
DSM / Lab IA



lea.berthomier@meteo.fr

Sommaire

- Formations
- Ressources de calcul
- Ressources documentaires
- Espaces de discussion
- Veille technologique
- Modèles de langage



Formations

1. Formation permanente :

Apprentissage Machine pour les curieux (3 jours)

Animée par Bruno Pradel (DSM/LabIA), 2 à 3 sessions par an.
Support accessible sur GitHub [ici](#).



2. En ligne :

FIDLE : Formation d'Introduction au Deep Learning

Créée par le CNRS, ouverte à tous, 20 séquences de 2h.
Toutes les séquences sont accessibles en différé sur Youtube [ici](#).

3. Formation à la demande :

Faire de l'apprentissage profond à MF en Python (2h)

Animée par le DSM/LabIA, 1 session par mois.
Pour les personnes qui veulent passer à la pratique.



Ressources de calcul

Possibilité d'accès à différentes machines avec GPUs :
 Bélénos, European Weather Cloud, Infra GPU du LabIA, SXBigData, Jean Zay, ...

Page Confluence pour vous diriger vers la machine adaptée à votre besoin : [lien](#)

Au besoin, contactez **DSM/LabIA** ou **SIRES**

1. Quelle machine avec GPU(s) pour mon projet IA ?

Mes critères	EWC	Bélénos	Infra Lab IA	Infra SIRES (sxbigdata)	Explications
Je dois accéder à de gros volumes de données (>10To) sur les bases de données Météo-France.	Red	Green	Orange	Orange	Le scratch du LabIA fait 100 To et est occupé à 95% (nettoyage prévu en Novembre 2023). Sur le EWC il faut creuser avec le CR (Paul Philis) et les admins des "Tenants" Météo pour voir quel "gros volume" serait allouable pour un projet.
Mon projet va durer moins de 6 mois. (ex. stage)	Green	Orange	Orange	Green	Pour un stage ou un projet court, le EWC permet de provisionner puis détruire une machine dédiée, avec un ticket d'entrée moindre que pour Belenos et le LabIA.
J'ai besoin de plusieurs GPU en parallèle pour une seule expérience IA.	Orange	Green	Green	Red	Les noeuds de Belenos et du LabIA sont multi-GPUs. Sur le EWC, chaque noeud est mono-GPU donc il est possible de faire du multi-noeud pour faire du multi-GPUs sur EWC mais avec un surcout en terme de performances.
J'ai besoin de lancer plusieurs expérience IA en parallèle.	Green	Green	Green	Green	Possible partout.
Je souhaite utiliser Docker pour gérer mon environnement de travail Python.	Green	Red	Green	Red	Bélénos ne permet pas de faire du conteneur pour le moment. (à creuser avec la DSI)
Je souhaite utiliser Conda pour gérer mon environnement de travail Python.	Green	Green	Orange	Green	L'infra du LabIA propose micromamba à titre expérimental uniquement.
J'ai besoin d'une grande bande passante entre le GPU et les données. (ex. chaque échantillon d'entraînement est volumineux)	Orange	Green	Orange	Green	L'infra du LabIA est limitée empiriquement à 1.5 Go/s entre le scratch et les noeuds GPUs, sur le EWC nos benchmarks indiquent 1.6 Go/s sur un noeud avec un disque supplémentaire attaché à la VM et monté sur /scratch.

Ressources documentaires

2 portails Confluence accessibles à tous

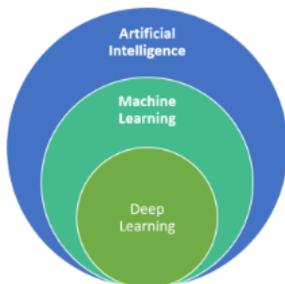
Lien portail DIAMS



DIAMS est le centre de Documentation et ressources en IA pour la Météo et les Services

Il fournit des informations générales sur l'IA appliquée à la météo et au climat, et organise l'accès à des ressources documentaires complémentaires et facilités de différents services et entités de Météo-France investies dans ces sujets.

Qu'est ce que l'intelligence artificielle ?



Que fait Météo-France dans le domaine de l'IA pour la météo et le climat ?

Quelques questions sur l'IA en Météo

Lien Centre de Ressources IA



Centre de ressources en IA

Se former au Deep Learning

- (LabIA) Rencontres R&D 2023: Du neurone humain à Pangu-Weather: petit plongeon ... par la pratique - vidéo d'une heure
- (Club Deep Learning) Inventaire des formations : à Météo-France, en ligne, pour les agents, pour les directeurs
- (GMAP/PREV) Formations et bibliographie générale (deep-learning général + appliqué à la PN)
- Support de Formation : Comment faire de l'IA en pratique à MF ?
- Veille technologique : comment se tenir au courant des avancées en IA ?

Notes techniques

Architecture Half-Unet 128 - Synthèse
Pandas: une fausse bonne idée
Sérialisation & Dé-sérialisation d'un modèle DL : une histoire de formats
Optimisation du calcul des métriques métiers : passage sur GPU
Le Receptive Field : késako ?

Méthodologie pour le développement en IA

1. Tuto Git et Gitlab
 2. Configurer son repo git pour un projet IA
 3. Comment VSCode va changer ta vie, en 10 étapes
- Créer un environnement Conda sur HPC
Créer un environnement Micromamba sur l'infra du Lab
- Inventaire des outils pour les projets IA/Deep-Learning
- Méthodologie pour réaliser un "gros" projet de Deep Learning
- Mise en OPER d'un projet IA - Bonnes pratiques pour les projets d'IA

Liens utiles

(DSM/GOA) Catalogue : Les bases de données : descriptions et utilisations

Utilisation des GPU

1. Quelle machine avec GPU(s) pour mon projet IA ?
 2. Guide utilisation des machines European Weather Cloud (EWC)
 3. Guide utilisation de l'infra GPU du Lab IA Administration de l'European Weather Cloud
- MAJ infra GPU du Lab IA en 2023

Les développements IA à MF

--> Le monorepo avec tous les codes du Lab IA

Inventaire des cadres de développement en apprentissage profond appliqué à la Météo

Pangu-Weather : déploiement à MF

Espaces de discussion

Club Deep Learning

Espace de discussion mensuel

Réunion tous les 2^e jeudi du mois, en présentiel et en virtuel

Réunion des pratiquants et curieux en IA, présentations sur la base du volontariat, sur les **travaux en cours, tuto, bug, bonnes pratiques...**



Bienvenue au Club Deep Learning

Vous souhaitez découvrir les projets IA en cours dans tout l'établissement ? Rejoignez le Club Deep Learning ! Ce séminaire mensuel et informel vous permettra de vous tenir informé.e des derniers développements. Pour recevoir les invitations au Club, inscrivez-vous à la liste de diffusion deep_learning@meteo.fr grâce à ce lien.

Le prochain Club Deep-Learning se réunira le  **7 mars 2024** à 14h30.

Inscription liste de diffusion :
deep_learning@meteo.fr

G3TIA

Coordination des agents du CNRM autour de l'IA.

Circulation d'information, interface avec l'extérieur du labo, montage de projets communs, auto-formation, contacts pour thèses, stages, organisation de séminaires...

G3T Intelligence Artificielle

Créée par BOULLOT Nathalie, dernière modification par ZAMO Michael le janv. 22, 2024

Groupe de Travail Thématique Transverse (G3T) sur l'intelligence artificielle et le machine learning

Actualités

Prochaine réunion : Un séminaire commun G3T-IA & Club Deep Learning est organisé le 8 Février 2024 au CIC et en visioconférence. Le site de l'événement est disponible avec le programme provisoire de la journée : <http://www.meteo.fr/cic/meetings/2024/JIA/> . Prière de vous inscrire sur ce site avant le 2 février 2024, pour des raisons de logistique (pauses café...).

Veille technologique

Le Machine Learning est un domaine qui évolue très vite, voire trop vite...

Quelques pistes pour rester à jour :

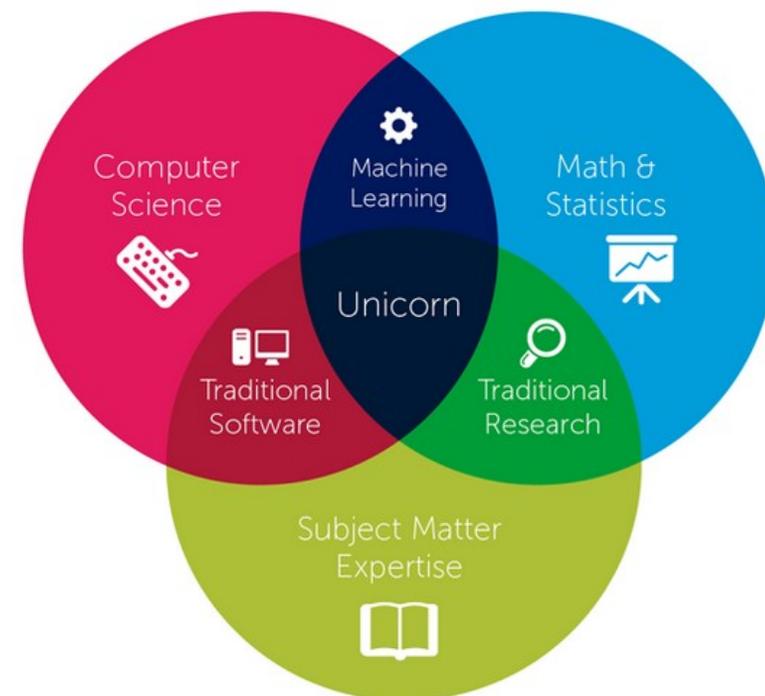
- Participer au **Club Deep Learning**
- S'abonner au **canal Rocket Chat #IA_info**
- **Listes de diffusion externes** : Data Elixir, Alpha Signal
- Lire des **blogs** : Statquest, Distill, Lil'Log, Google Research

Ne pas négliger l'aspect **développement informatique** !

- Canal RocketChat #python
- Liste de diffusion Pycoder's Weekly
- Blog pythonspeed

Tous les liens accessibles [ici](#)

Data Science



Modèles de langage



ChatGPT



Bard

Gemini



Bing



GitHub
Copilot

LLM : Large Language Models

Ces outils peuvent servir à traduire du texte, générer du code, corriger la grammaire ou l'orthographe, rédiger un compte-rendu, expliquer un concept, etc.

Attention aux fuites de données ! Toutes les données que vous envoyez à travers ces services deviennent la propriété d'entreprises américaines, et pourront être ré-utilisées pour entraîner les modèles.

Page de présentation sur **Intramet sur l'IA Générative** par Abdelkader Beldjilali : [lien](#)