

NLP pour la sensibilisation aux problèmes du vivant



GreenAI
U · P · P · A



Qui sommes-nous ?

1

design new low consumption AI based on mathematical results and algorithms,

3

build a local ecosystem around AI for environment: new research, new products, new training and partnerships.

2

apply AI to solve the next environment challenges,

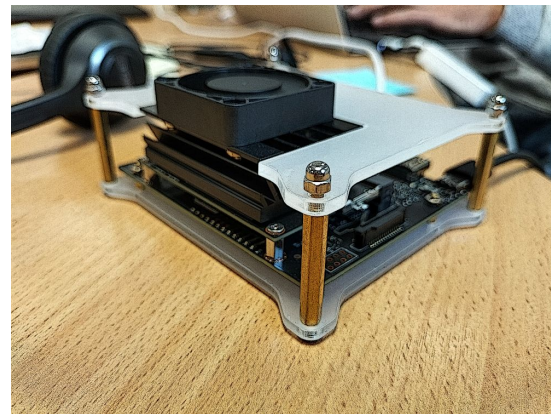


1 Power-efficient Machine Learning

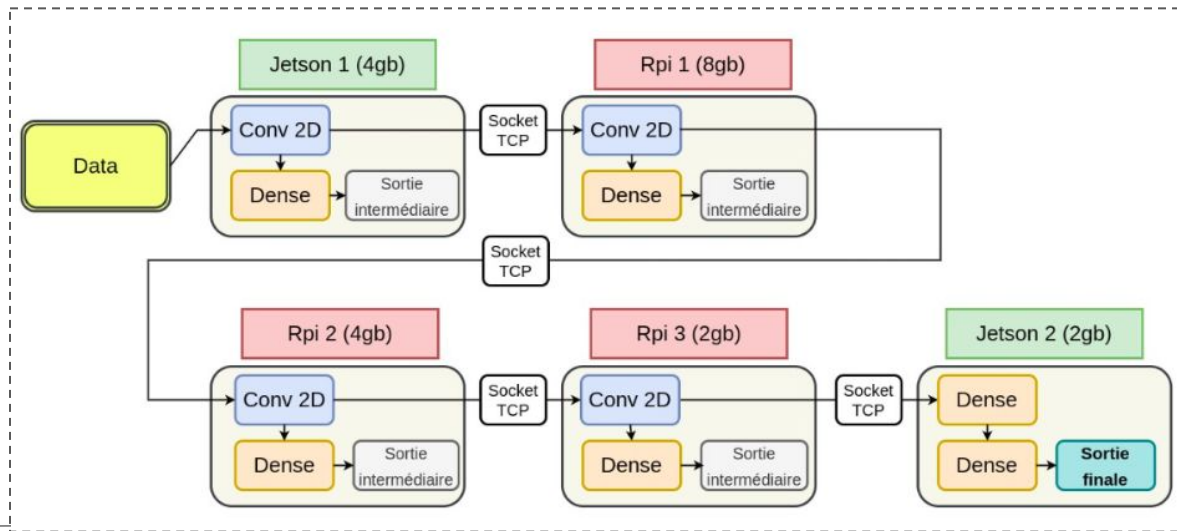
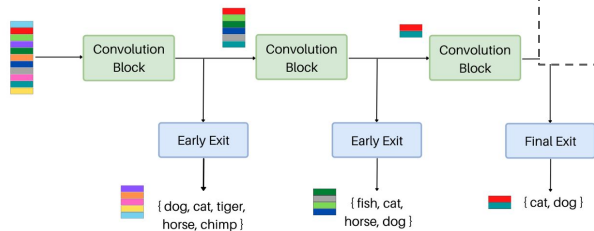
- Maths stats based deep/machine learning
- Low cost algorithms
- Decentralized and Edge deep/machine learning
- Comprehensive carbon footprint of data centers

Objectives and deliverables:

Theoretical foundations of new algorithms, publications and academic communication and workshops, open source softwares, prototypes, training ...



1 New hardware



Raspberry pi

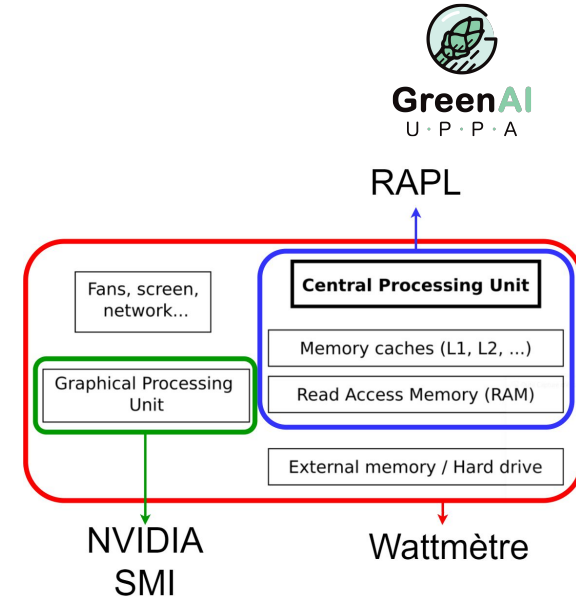
1 Measurement tools

- Hardware
 - Connected power meters
- Softwares
 - Library based on RAPL and nvidia-smi
 - [AIPowerMeter](#)
- Fine grained measurement of AI algorithms
- Other application : measure of consumption at the scale of a data center
 - LabIA : 12 stations dedicated to AI used by 5 laboratories
 - Analysis of data scientists usage

<https://greenai-uppa.github.io/Coca4AI/>

Objectives:

Opensource software, apply AI to real environmental applications





GreenAI

U · P · P · A



Prof en Poche



GreenAI

U · P · P · A



Prof en Poche

HIZKIA
INFORMATIQUE



GreenAI
U · P · P · A



Prof en Poche

HIZKIA
INFORMATIQUE





GreenAI
U · P · P · A



Prof en Poche

HIZKIA
INFORMATIQUE



EMAC

Technical Rubber Compounds

Le projet du DATA CHALLENGE



- Base de données
- Interface Web Django : jeu, moteur de recherche, interface ADMIN

Objectif : sensibiliser notre communauté aux problèmes du vivant

Objectifs du data challenge

- **Améliorer le modèle NLP** (aujourd'hui fasttext entraîné sur 300k tweets) pour deux raisons :
 - (1) avoir un moteur de recherche plus pertinent
 - (2) avoir une aide à l'annotation (nouveau tag, nouveau contenu)

Objectifs du data challenge

- **Améliorer le modèle NLP** (aujourd'hui fasttext entraîné sur 300k tweets) pour deux raisons :
 - (1) avoir un moteur de recherche plus pertinent
 - (2) avoir une aide à l'annotation (nouveau tag, nouveau contenu)
- **Intégrer ces améliorations** dans l'interface python django

Objectifs du data challenge

- **Améliorer le modèle NLP** (aujourd'hui fasttext entraîné sur 300k tweets) pour deux raisons :
 - (1) avoir un moteur de recherche plus pertinent
 - (2) avoir une aide à l'annotation (nouveau tag, nouveau contenu)
- **Intégrer ces améliorations** dans l'interface python django
- **Rejoindre l'équipe** d'annotateurs et de sensibilisateurs

Les données

- 981 contenus annotés (30 tags au total) au format json
- 38Mo = 308 661 tweets sur l'environnement (communauté @Bonpote)
- Autres communautés disponibles (200Mo de tweets potentiels)

Les données

- 981 contenus annotés (30 tags au total) au format json
- 38Mo = 308 661 tweets sur l'environnement (communauté @Bonpote)
- Autres communautés disponibles (200Mo de tweets potentiels)

Pour l'évaluation :

- 20 requêtes du moteur, avec 20 réponses idéales
- 20 requêtes du moteur, avec 20 réponses idéales

Livrables



Modèles de NLP entraînés + notebooks
python démontrant la pertinence du modèles

Livrables



Modèles de NLP entraînés + notebooks
python démontrant la pertinence du modèles



Ajout de fonctionnalités à l'interface
python-django (front + back)

Livrables



Modèles de NLP entraînés + notebooks
python démontrant la pertinence du modèles



Ajout de fonctionnalités à l'interface
python-django (front + back)



Un show final pour le jury

Le code

Disponible sur [Github ici](#) et contient :

- l'arborescence django du site contenant toute la machinerie (models, views, templates), le moteur de recherche actuel, les jeux, etc
- un starter-pack contenant les données, et un script python pour l'évaluation des 20 requêtes de tests

Conseil pour débuter

⚠ Attention : Moteur de recherche + classification

👨‍⚖ Penser intégration dès le début : [tutoriel django](#)

🤝 Répartissez-vous bien les rôles



Merci !



Sébastien Loustau - Porteur du projet 0783273883

sebastien.loustau@univ-pau.fr