

# Marc ALPHONSUS

## Data Science, +8 ans d'XP

Data scientist senior avec un large éventail d'expériences; du scoring à la détection de fraude, en passant par la reconnaissance d'images et la modélisation du langage dans des contextes big data. Je recherche un poste de lead tech data ou data scientist senior. Versé au travail en équipe, imprégné de la philosophie devops, j'apporte beaucoup d'attention à l'industrialisation et l'automatisation des modèles.

Mes domaines d'expertise tournent autour du Deep Learning en particulier le traitement du langage naturel et l'apprentissage par renforcement.

Niort, Nouvelle Aquitaine  
31 ans, nationalité Française  
Permis A, B

alphonsus.marc@gmail.com  
[linkedin](#)  / [github](#)   
+33 6 62 62 62 48

---

### Expérience

#### CGI Lead Data Science 2018 - 2020

Manager le lab IA CGI en offrant une vitrine des savoirs-faire Data Science. Construire et valoriser la stratégie IA des clients Grand Ouest.

#### MACIF :

- Macif NLP: Création des modèles de traitement des flux texte pour le groupe (routage/classification, extraction d'entités - i.e. NER, analyse du sentiment)- amélioration de la performance de 85% sur 3 classes à **92% sur 56 classes** > [tensorflow](#), [huggingface](#), [spacy](#), [docker](#)
- Hub IA: Exposition d'une plateforme de modèles d'IA sécurisée et monitorée utilisables par tous les salariés du groupe sous forme d'un catalogue swagger d'API > [WSGI](#), [API Gateway](#), [docker](#)
- AutoML Workflow : Conception d'un pipeline de réentraînement automatique des modèles en production - Les modèles sont auto-alimentés grâce à une boucle de feedback des utilisateurs et réentraînés de manière spontanée mensuellement (active learning, multitask-learning) > [python](#), [gitlab CI](#), [docker](#)
- RAD/LAD: Mise en place d'un service automatique d'analyse de documents (classification automatique et OCR) avant GED **95% sur 80 classes** > [pytorch](#), [docker](#)
- MeetUP: Définition d'une stratégie data à l'échelle du département et évangelisation des concepts clés de la data science auprès des différentes directions du groupe [OKR](#), [Gslides](#), [Jupyter](#)
- Chatbot EPI: refonte l'UX de l'espace personnel internaute - la mise en place de l'IA a permis de tenir les objectifs clés de l'entreprise (délai de traitement d'une demande <1 jour) avec une meilleure qualification du besoin (extraction du produit, du bénéficiaire, etc) > [tensorflow](#), [spacy](#), [rasa](#)

#### IMA (InterMutuelles Assistance) :

- Assistance Augmentée: PoC d'une IA d'assistance téléphonique - Combinaison d'un modèle d'ASR/STT et de modèles de traitements du langage (NER, Classification, Similarité) pour assister et accélérer l'assistance lors des pannes et accidents de la route;  
Process mis en place : Retranscription de la conversation, identification du véhicule, géolocalisation du

client, compréhension de la panne à partir de la description du client, délégation d'un dépanneur/fourniture de service (taxi, hôtel...) et envoi d'un SMS récapitulatif > [pytorch](#), [Go](#) , [docker](#)

#### CGI - Lab IA Grand Ouest :

- Management et mentorat de collaborateurs
- Avant-vente, cadrage/gestion de projet
- Veille technologique et mise en place des learning days
- Conférences et vulgarisation

#### MAIF Data scientist 2014 - 2018

Apporter aux départements porteurs de projets, l'expertise nécessaire en apprentissage statistique et concevoir des outils de demain au service du sociétaire.

- Mail: Routage automatisé des e-mails et résumé automatique, devenu [Mélusine](#) ; design de l'architecture des réseaux neuronaux et des algorithmes de résumés extractifs > [tensorflow](#), [scikit-learn](#), [docker](#)
- MAIF & Go: Tarification d'une offre automobile connectée à partir des comportements de conduite remontés via capteurs GPS (IoT) > [pySpark](#), [R](#), [caret](#)
- Valeur Sociétaire: Modélisation de la rentabilité des clients et prospects à horizon long (LTV) > [pySpark](#), [Hadoop Map/Reduce](#)
- Téléphonie: Prédiction de la charge d'appels pour dimensionnement RH des plateaux téléphoniques **taux d'erreur à 5%, proposition de publication** > [R](#), [caret](#)
- Détection d'anomalies: Identification de risques sous-assurés (fraude & devoir de conseil) > [scikit-learn](#)

#### IRIAF Chargé de TD Machine Learning 2017 - 2019

Initier des étudiants à l'apprentissage automatique - programme: théorie de la décision, régression logistique, SVM, CART, forêts aléatoires et réseaux de neurones simples et convolutionnels.

#### MAPA Statisticien 2013 - 2014

Concevoir et réaliser des études visant à expliquer et corriger les dérives tarifaires du portefeuille, l'élasticité au prix des sociétaires et les résultats financiers.

---

### Compétences techniques et expériences extra-professionnelles

#### Projets choisis

**TextNets:** Mes modèles autour du NLP, répliques d'articles, analyse de similarité textuelle, résumé automatique, traduction, search...

**DeepCount:** dénombrement sur image fortement peuplées, estimation par density map

#### Hackathons

**Hackathon IMA 2019:** Première place - Hackathon data

**AfLPC x LineUp7:** Seconde place - Tech For Good

**Languages:** Python, R, SAS, Go

**Langues:** English, Français, Deutsch

**Frameworks:** tensorflow, pytorch, scikit-learn, Apache Spark, opengym

**Cloud:** AWS, Google Cloud Platform

**Basics:** UNIX, algorithms, git, docker, redis, MongoDB, SQL

---

## 🎓 Diplômes & MOOCs

<b>2019</b>	<b>Spinning up Reinforcement Learning;</b> OpenAI <b>6.S094;</b> MIT
<b>2018</b>	<b>CS224n, CS229, CS231n;</b> Stanford University
<b>2017</b>	<b>Deep Learning Specialization;</b> Deeplearning.ai x Coursera
<b>2014</b>	<b>Machine Learning;</b> Stanford University x Coursera <b>Deep Learning;</b> Google x Udacity
<b>2013</b>	<b>Master, Statistiques Appliquées;</b> Université de Poitiers
<b>2010</b>	<b>Licence, Mathématiques Informatique et Applications;</b> Université Paris V