Samah GHALLOUSSI

65 boulevard de Sébastopol
75001 PARIS
Tel : 07.69.23.65.43
Email : samah.ghalloussi@gmail.com

Data Scientist NLP
Entrepreuneure d'Intéret Général (EIG2)
CEO/CTO AALIA.TECH
Anglais : *niveau avancé 920 / 990 au TOEIC*

¤ Formation Académique

Juin 2018	**Data Science Summer School 2018 – Ecole Polytechnique, Palaiseau**
Sept 2009 - 2013	**Ecole Polytechnique de l'Université Joseph Fourier, Grenoble I** Diplôme d'Ingénieur, spécialité Technologies de l'Information pour la Santé
Sept 2007 - 2009	**Faculté des Sciences de Luminy, Aix-Marseille Université** 1ère et 2e année Licence Biologie Informatique Mathématique, Major de Promo

¤ Domaines de Compétences

Langages : Python, C++, DART, C, C#, JAVA, UML, XML, R
Bases de Données : MongoDB, mySQL, PostgreSQL, Visual Basic, Access
Traitement Automatique des Langues : NLTK, Gensim, SpaCy, BERT
Machine Learning : Scikit-Learn, TensorFlow, Keras

¤ Expériences Professionnelles

Startup **Aalia Tech** 02/2019 – Aujourd'hui	**CEO/CTO & Data scientist** Contexte : Assistant vocal interprète médical pour les professionnels de santé Outils techniques : Python, Flutter, API, Gitlab, Cloudrun, Heroku Réalisations : - Collecte et normalisation des données - Modélisation de la base de données - Création d'un algorithme permettant la traduction entre patient et médecin en prenant en compte le contexte médical et culturel du patient - Gestion et animation d'ateliers de travail avec les Key Users (professionnels de santé : médecins, infirmiers, sage-femmes, …) - Appel à un webhook sécurisé contenant le service développé en Python et hébergé via Cloudrun - Création de plusieurs web services - Développement de l'application mobile dédiée à l'utilisation du système
M2I Formation 05/2022 – Aujourd'hui	**Formatrice Algorithmique, Sécurité des applications**
EPSI 02/2020 – Aujourd'hui	**Intervenante à l'EPSI (Ecole d'ingénierie informatique)** Contexte : Cours de programmation, machine learning pour les Bac+3 à Bac+5 Outils techniques : Tinkercard, Python, Django
Air Liquide 11/2021 - 12/2021	**Mission Freelance - Expert Modélisation base de données, SQL** Contexte : Mise en place et migration de la base de données Outils techniques : SQL Server
Vivadata 04/2019 – 05/2019	**Intervenante formation Machine Learning - Vivadata** Contexte : Tutorat/mentorat pour des cours de machine learning Outils techniques : Python, Anaconda, Scikit-Learn, Github, Jupyter
Ministère des Solidarités et de la Santé, DREES 03/2018 – 01/2019	**Entrepreuneure d'Interêt Général, 2e Promotion** Contexte : Valorisation des données du Système National de Données de Santé Outils techniques : Python, Spyder, Scikit-Learn, Github, Gitlab, Jupyter, SPARQL Réalisations : - Valorisation des données du SNDS (Système National des Données de Santé) - Génération automatique de rapport au format docx qui prend en compte le cahier des charges et les specs d'études statistiques. - Extraction d'information de tableaux issus de pdf : mise en place d'un système de classification supervisée multiclasse : algorithme utilisé SVM - Collaboration avec l'ANSM pour le développement d'un outil de veille sanitaire
Startup **Stryng Messaging Inc** Paris 06/2017 – 03/2018	**Ingénieur R&D** Contexte : Application de messagerie multisupport pour les professionnels (ventes automobiles) Outils techniques : Linux, Github, MongoDB, Python, Scikit-Learn, NLTK, AWS, Jira Réalisations : - Rédaction des spécifications technico-fonctionnelles détaillées, modélisation et gestion du Dataset (attributs, messages clients, …) - Création d'un modèle de prédiction pour la détection des futurs acheteurs - Service d'extraction d'entités nommées à partir de signatures d'emails pour le remplissage automatique du CRM - Travaux sur le résumé automatique à partir des conversations téléphoniques entre le vendeur et son client : utilisation de plusieurs algorithmes de machine learning (SVM, réseaux de neurones, - Service de tri d'emails de boites de messagerie : Utilisation d'un algorithme de classification non-supervisée (Spectral Clustering) pour classifier les sujets d'emails par centres d'interêts.
Commissariat à l'Energie Atomique CEA Saclay 06/2014 – 06/2017	**Ingénieur-Chercheur NLP** Contexte : Département Intelligence Ambiante et Systèmes Interactifs (DIASI) Laboratoire Vision et Ingénierie des Contenus (LVIC) Outils techniques : Linux, Kdevelop, Python, PostgreSQL, Github, Gitlab, WEKA, Scikit-Learn, NLTK, C++ Réalisations : - Création d'un outil de veille réglementaire pour Bureau Veritas avec filtrage et classification des données : algorithme SVM (Machine à vecteur de support) - Création de modèle de prédiction pour la classification de très larges datasets - Developpement d'un outil en C++ d'extraction de relation avec algorithme CRF (Conditional Random Field) - Développement d'un outil de recherche d'information multilingue optimisé via algorithme Spectral Clustering - Création d'agents conversationnels avec diverses plateformes : Dialogflow, Watson Conversation IBM, Wit.ai de Facebook et LUIS de Microsoft