

# Nabil Cheikh

*Ingénieur Data Scientist*

17 Rue du Marché  
92000 Nanterre, France

☎ 06 60 77 37 42

✉ [nabil.cheikh13@hotmail.com](mailto:nabil.cheikh13@hotmail.com)



## Compétences

Langages informatiques Python, R, Java, MySQL, Shell/Bash, C/C++, C#

Web Javascript, Typescript, Angular, React, HTML/CSS, Selenium, PHP

Librairies Python Scikitlearn, Numpy, Pytorch, Pandas, Matplotlib/seaborn, Tensorflow, Keras, FastAPI, Uvicorn, Langchain, Pydantic

IDE Visual Studio Code, Pycharm, Jupyter, IntelliJ, DBeaver

Outils pour développeurs Docker, AWS, Google Cloud Platform, (Putty/Filezilla/WinSCP), Jira, Jenkins, Git, Postman, Gitlab, Google Colab

## Expérience professionnelle

2023-2024 **Consultant développement fullstack et IA, CDI**, Daveo, Paris.

### Tâches:

- Création d'un site web utilisant un chatbot :
  - Utilisation du framework React pour la partie front-end : user interface similaire à celle de chatGPT, génération des messages du bot en streaming, possibilité de créer de nouvelles discussions
  - Utilisation de Python et des librairies FastAPI Uvicorn pour la partie back-end : utilisation de la librairie boto3 pour interagir avec les services AWS. Utilisation du modèle Anthropic Claude pour la génération de texte, proposé par le service AWS Bedrock.
  - Customisation du chatbot via la librairie python Langchain afin de fournir au modèle une user story : le contexte selon lequel le bot doit interagir avec l'utilisateur (ici, c'est un assistant), ainsi que le contexte de données (ici, c'est les données internes de la société Daveo pour leurs employés : comment poser des congés, qui est le PDG, etc)
  - Liaison front-end back-end via requêtes d'API et gestion des réponses obtenues ainsi que débogage.
- Préparation aux certifications AWS

2022-2023 **Ingénieur développement**, *CDI*, ArianeGroup (prestataire Astek), Les Mureaux.

**Tâches:**

- Conception et implémentation de pipelines de preprocessing de données en Python, incluant des tâches de feature engineering et de cleaning, dans le cadre de la préparation des données pour la simulation de lancement de fusées.
- Mise en place de pipelines pour l'entraînement de modèles de machine learning visant à détecter des anomalies dans les scénarios de décollage.
- Développement d'algorithmes de traitement des données pour l'optimisation des performances des simulations et la détection précoce d'anomalies critiques dans les étapes de décollage.
- Automatisation des processus de gestion des données via des scripts Bash/Shell pour la collecte, le nettoyage et l'agrégation de données depuis différentes sources.
- Intégration des résultats dans un tableau de bord de surveillance en temps réel des anomalies pour l'équipe d'ingénierie.
- Gestion des outils DevOps : Jenkins, Jira, SVN, FishEye pour l'intégration continue et la collaboration entre équipes.
- Autre : Rédaction de spécifications techniques et fonctionnelles, mise en production des systèmes d'analyse de données.

2020-2022 **Développeur fullstack**, *CDI*, Acpqualife, Aix-en-Provence.

**Tâches:**

- Développement frontend : Enrichissement d'un progiciel par ajout et modifications d'écrans via Angular, correction d'anomalies, élaboration de suites de tests unitaires automatisés
- Développement backend : Implémentation de scripts automatiques en Spring Batch (batchs), correction d'anomalies, gestion de webservices
- Initiation DevOps : Formation sur Docker, gestion des outils Jenkins et Jboss pour les processus de build automatiques et de déploiement, livraison mensuelle aux clients des résultats des campagnes de tests automatiques
- Autres : Rédaction de spécifications fonctionnelles et techniques

## Projets

### 2024 **Data science**, *Londres*, Hackathon - Mistral AI.

- Participation au Hackathon de MistralAI organisé par a16z dans l'espace événementiel CodeNode. L'objectif étant de développer une app intégrant les outils IA de MistralAI en un week-end
- Projet choisi : Algorithme de détection de potentielles attaques de prompt injection ciblant les sites web intégrant un outil LLM customisé.
- Contexte : un client souhaitant avoir un chatbot ou un outil LLM, intégré dans son site web. Il peut le customiser en lui ajoutant un context prompt : il lui définit son rôle lors de l'interaction avec un utilisateur (Exemple : "Tu es un assistant IA spécialiste en analyse du marché financier").
- Des utilisateurs malicieux peuvent interagir avec le bot de sorte à briser son prompt de base et fournir des données sensibles le faisant sortir de son rôle défini.
- Solution : améliorer le context prompt afin de le rendre plus robustes face aux attaques de prompt injection.
- Prompt engineering combiné au tree of thoughts (ToT) prompting :
  - ToT du modèle LLM ayant pour rôle d'attaquer le context prompt, fournissant les meilleures attaques
  - ToT du modèle LLM ayant pour rôle de contre-attaquer ces meilleurs attaques en fournissant de nouveaux context prompt plus robustes.
  - Répétition de ce processus jusqu'à ce que toutes les attaques contre le context prompt échouent.

### 2024 **Data science**, *Paris*, Projet Le Wagon.

- Création d'une IA capable de jouer à Tetris à l'aide du Reinforcement Learning en utilisant l'algorithme de Deep Q-learning
- Utilisation de la librairie Pyboy pour émuler le jeu via un fichier ROM de Tetris
- Construction d'un modèle Deep learning ayant en entrée l'état du jeu (nombre de trous, variation de hauteurs, hauteur max, pièce à poser), et donnant en sortie la suite d'actions à effectuer pour passer à l'état suivant
- Stockage des données du jeu destinée à la mémoire de l'agent qui s'en servira pour améliorer sa prise de décision : chaque ligne de la mémoire est une combinaison de données.
  - Donnée 1 : État initial
  - Donnée 2 : Prochain état
  - Donnée 3 : Récompense obtenue
  - Donnée 4 : Partie terminée ou non (booléen)
- MLOps : gestion des services GCP pour un déploiement plus efficace de l'application :
  - Gcloud Instance pour entraîner le modèle en background dans une VM sur plusieurs jours
  - Gcloud Bucket pour gérer automatiquement la sauvegarde du modèle entraîné, avec une gestion de checkpoints afin de pouvoir mettre en pause l'apprentissage du modèle et le reprendre quand on le souhaite

---

## Education

2024–2024 **Data Science & AI course | Le Wagon , Paris.**

**Apprentissage:**

- Data science et machine learning:
  - Apprentissage de différents modèles avec scikit-learn
  - Pipelines pour pré-traitement des données
  - Compression d'images, algorithmes de détection de spam, modèle de prédiction pour les prix de l'immobilier, délais de livraison, prix d'une action etc
  - Modèles complexes : Decisiontree, Randomforest, boosting de gradient et plusieurs optimiseurs
  - Apprentissage non supervisé : PCA, Kmeans
  - Compréhension des techniques du NLP
- Deep Learning:
  - Création de réseau neuronal avec tensorflow keras
  - Analyse under the hood : gradient loss, optimiseurs, backpropagation, régularisation, vanishing gradient, momentum
  - CNN pour la détection d'objets sur les images, MaxPool2D, Data Augmentation
  - RNN comme apprentissage profond pour prédire l'image suivante ou le mot suivant d'une phrase, NLP appliqué à l'apprentissage profond, création de transformateurs à partir de zéro, attention

2017–2021 **École d'ingénieur Polytech , Marseille.**

**Apprentissage:**

- Programmation impérative et orientée objet, algorithmique, architecture des ordinateurs et réseaux, sécurité informatique, web, intelligence artificielle et jeux
- Génie logiciel, architecture des applications d'entreprise (JEE), sécurité des systèmes d'information, modélisation et conception objet, systèmes d'exploitation, applications web, mobiles et multimédia, bases de données SQL/NoSQL
- Approfondissement dans les systèmes d'informations : maîtrise des technologies de traitement d'informations et des connaissances (IA, datascience, cloud), sécurité des cloud, urbanisation des SI, business intelligence

2015–2017 **Parcours des Ecoles d'Ingénieurs Polytech (PeiP), Marseille.**

**Apprentissage:** Mathématiques (Algèbre, Analyse, Géométrie), Physique (Physique newtonienne, électrocinétique, thermodynamique, électromagnétisme, électronique numérique, mécanique, ondes et signaux), Ingénierie (Initiation aux outils informatiques, analyse fonctionnelle, algorithmique et programmation python, outils pour le web)

---

## Langues

Français - Langue maternelle

Anglais - Bonne maîtrise, TOEIC 810pts

Italien - Maîtrise scolaire

---

## Hobbies

Piano, Escalade, Musculation