

## Sujet de Stage de Master

### Etude de la Généralisation d'une approche de Machine Learning pour l'analyse de défaillance sur des systèmes micro-électroniques.

**Domaine Scientifique :** Machine/Statistical Learning, Analyse de défaillance

**Laboratoire :** UMR LIMOS -Ecole des Mines de Saint Etienne

**Encadrant:** Pr. Mireille Batton-Hubert ([batton@emse.fr](mailto:batton@emse.fr)); Pr. Xavier Boucher ([boucher@emse.fr](mailto:boucher@emse.fr)) et Dr. Anis Hoayek ([anis.hoayek@emse.fr](mailto:anis.hoayek@emse.fr))

**Rémunération:** Allocation de stage (environ 580 €/mois)

**Mots clés :** Deep Machine Learning, Auto-Encoder / Variational Auto-Encoder, Modélisation des données textuelles (NLP), Techniques de vectorisation, Analyse de défaillance, Pre-trained Transformer models

### Objectifs

En analyse de défaillance des systèmes micro-électroniques, les experts du domaine, sachant la description de la défaillance, proposent une série d'analyses à faire afin d'arriver à une cause fondamentale de la défaillance. Toute cette démarche est stockée sous forme de différentes variables dans une base de données majoritairement textuelle. Le travail sur telle base de données a été initié par la thèse de Kenneth Ifeanyi EZUKWOKE qui a développé un outil d'aide à la décision couvrant plusieurs aspects : 1) Prétraitement et vectorisation des données textuelles à travers plusieurs techniques de NLP ; 2) Réduction de dimension et classification de l'espace sous-jacent en appliquant des techniques basées sur des réseaux de neurones (e.g., Auto-Encoder) couplées avec des méthodes de clustering (e.g., Gaussian Mixture Model (GMM)) ; 3) Adapter un pre-trained model de type transformer (e.g., GPT2/3) pour être capable, suite à l'obtention d'une description de défaillance, de générer le chemin optimal d'analyses à faire par un expert pour aboutir à la cause fondamentale de la défaillance.

Ainsi, l'objectif du Master est de consolider la généralisation des méthodes de Machine Learning/Statistical Learning développés dans le cadre de la thèse de Kenneth Ifeanyi EZUKWOKE, sur d'autres cas tests issus de jeux de données fournis par les industriels du consortium FA 4.0. D'un point de vue académique il s'agira d'identifier à partir des données, les structures et caractéristiques intrinsèques - statistiques qui permettront d'évaluer la généralisation des méthodes développées à partir de différentes techniques issues des sciences de données adaptées aux données textuelles et de leur espace de représentation.

La mise en place de ces tests permettront d'explorer l'impact de différents contenus sur la performance des algorithmes mis au point dans le cadre des travaux de la thèse de Kenneth Ifeanyi EZUKWOKE. De plus, ils sont un enjeu important pour les industriels du consortium afin d'identifier également des différences de pratiques industrielles dans le mode d'analyse des défaillances et constitue un élément important pour la bonne finalisation des missions assurées par Mines Saint-Etienne dans ce projet.

### Profil du Candidat

Etudiant en Master dans le domaine de science des données, de machine learning et intelligence artificielle, du génie

Certaines des compétences suivantes seront particulièrement appréciées pour le projet :

- Modèles de deep Machine Learning (réseaux de neurones, Auto-Encodeur)
- Analyse des données textuelles (les techniques de NLP)
- Modèles génératifs de type transformer (GPT2/3)
- Compétences avancées en programmation sous Python,

Un intérêt pour le secteur applicatif de micro-électronique serait apprécié.

### **Contact et candidature**

Les candidatures doivent inclure : Curriculum Vitae détaillé + lettre de motivation

Candidature à envoyer à l'un des encadrant du stage.