

**Post-doctorat en sciences de données : NLP, deep-clustering, classification et deep learning appliqués à l'analyse des défauts en production micro-électronique  
Institut Henri FAYOL**

**Type de poste :** Post-doctorat (candidat disposant d'une thèse). Les candidats 'ingénieur de recherche', non titulaires de doctorat mais avec une forte expérience dans le domaine peuvent nous contacter directement.

**Contrat :** Poste ouvert en CDD.

**Durée :** 18 mois

**Responsable Scientifique :** Pr. Mireille Batton-Hubert, EMSE/FAYOL - LIMOS.

**Responsable du Projet :** Pr. Xavier Boucher, EMSE – Institut FAYOL - LIMOS.

**Date de démarrage souhaité :** janvier 2022

**Collaboration industrielle et européenne :** STMicroelectronics (Grenoble Reliability & Analysis Laboratory), Projet Européen FA4.0 incluant 15 partenaires européens scientifiques et industriels.

**Mots clés :** NLP, Data-mining, Machine Learning, Deep Learning, Dimension reduction, Clustering, Classification, Industrie micro-électronique

**Presentation de l'Ecole**

L'École Nationale Supérieure des Mines de Saint-Étienne (Mines Saint-Etienne), École de l'Institut Mines Télécom, sous tutelle du Ministère de l'Économie, des Finances et de la Relance est chargée de missions de formation, de recherche et d'innovation, de transfert vers l'industrie et de culture scientifique, technique et industrielle.

Mines Saint-Etienne représente : 2 200 élèves-ingénieurs et chercheurs en formation, 400 personnels, un budget consolidé de 46 M€, 3 sites sur le campus de Saint-Étienne (Loire) d'environ 26 000 m<sup>2</sup>, le campus Georges Charpak Provence à Gardanne (Bouches-du-Rhône) d'environ 20 000 m<sup>2</sup>, 5 centres de formation et de recherche, 6 Unités de Recherche, un centre de culture scientifique technique et industrielle (La Rotonde). Mines Saint-Etienne a des projets de développement sur Lyon, notamment sur le Campus Numérique de la région Auvergne-Rhône-Alpes et de nombreuses collaborations à l'international. Le classement du Times Higher Education (THE), nous place en 2021 au niveau mondial dans la gamme 301–400 dans le domaine de l'Engineering (6<sup>ème</sup> école d'ingénieurs en France et 1<sup>er</sup> établissement dans ses deux régions d'appartenance) ainsi que dans les domaines Computer Science (501-600) et Physical Sciences (601-800).

Le Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes (LIMOS), qui sera le laboratoire d'accueil, est une Unité Mixte de Recherche (UMR 6158) en informatique, et plus généralement en Sciences et Technologies de l'Information et de la Communication (STIC). Le LIMOS est principalement rattaché à l'Institut des Sciences de l'Information et de leurs Interactions (INS2I) du CNRS et de façon secondaire à l'Institut des Sciences de l'Ingénierie et des Systèmes (INSIS). Il a pour tutelles académiques l'Université Clermont Auvergne et Mines Saint-Etienne (MSE). Le positionnement scientifique du LIMOS est centré autour de l'Informatique, la Modélisation et l'Optimisation des Systèmes Organisationnels et Vivants. Le poste est à pourvoir dans l'axe Modèles et Algorithmes de l'Aide à la Décision (MAAD) ou Systèmes d'information et de communication (SIC).

Créé en 2011, l'Institut Henri Fayol fédère l'ensemble de ses équipes d'enseignants chercheurs en mathématiques, génie industriel, informatique, environnement et en management autour des thématiques de l'efficacité, de la résilience et de la durabilité de l'industrie et des territoires du futur. Dans une perspective de développement durable et de responsabilité sociétale, l'efficacité et la résilience des entreprises et des territoires dans lesquels elles s'insèrent, doivent en effet être envisagées non seulement sur le plan technique et économique, mais aussi au niveau social, environnemental. Deux plateformes technologiques ont été développées pour valider, promouvoir et enseigner le travail effectué au sein de l'institut dans des conditions quasi réelles. La première est dédiée au territoire du futur (Plateforme Territoire <sup>1</sup>) et la seconde à l'industrie du futur (Plateforme IT'M Factory<sup>2</sup>). Au sein de cet institut, le département Génie Mathématique et Industriel (GMI) s'intéresse aux modèles de décision et aux sciences des données, abordés en particulier au travers de la probabilité, la statistique, l'optimisation et la recherche opérationnelle, liés aux méthodes quantitatives de résolution de problèmes qui sous-tendent la conception et l'optimisation de systèmes réels et complexes.

Plus particulièrement, les activités de recherche des enseignants chercheurs du département GMI (*Génie Mathématique et Industriel*) sont développées au sein du LIMOS. Elles sont centrées sur le développement de solutions mathématiques pour l'aide à la décision, comme la prévision, la modélisation et l'optimisation de systèmes continus ou procédés discrets. Ces domaines scientifiques sont ceux des probabilités, des statistiques, des Sciences de données, de la méta-modélisation pour l'estimation (thème Données, Services et Intelligence – DSI – de l'axe SIC), de la modélisation et de l'optimisation qu'elle soit continue (axe MAAD) ou discrète (axe ODPS du LIMOS), locale et ou globale. Les activités de recherche sont soutenues par de nombreuses collaborations scientifiques au travers d'outils phares, notamment de 3 chaires de recherche - chaire ValadoE<sup>3</sup> - chaire Corenstock<sup>4</sup> - consortium Ciroquo (ex chaire OQUAIDO<sup>5</sup>). Elles s'appuient sur des compétences mathématiques pour la résolution de problèmes autour notamment de la méta-modélisation, de l'optimisation, et de la théorie de la décision en lien avec la simulation et la prise en compte de l'incertitude. Il s'agit alors d'intégrer la diversité et la dimension de données disponibles, remontées au travers des différents services et/ou réseaux de capteurs. L'objet est de valoriser cette donnée et d'adapter les outils de prévision, de simulation et d'optimisation pour l'industrie notamment au service de la prise de décision ; cette dernière se décline souvent dans un cadre combinant à la fois des problématiques de systèmes continus et de process discrets.

### Contexte scientifique et industriel

Ce post-doctorat s'inscrit dans le cadre du projet européen FA4.0, en collaboration avec les équipes de STMicroelectronics sur Grenoble (Grenoble Reliability & Analysis Laboratory). Plus précisément, le projet vise

<sup>1</sup> <https://territoire.emse.fr/>

<sup>2</sup> <https://www.mines-stetienne.fr/entreprise/itm-factory/>

<sup>3</sup> <https://www.mines-stetienne.fr/recherche/5-centres-de-formation-et-de-recherche/institut-henri-fayol/actualites/>

<sup>4</sup> <https://www.mines-stetienne.fr/panoramines/2021/03/23/mines-saint-etienne-partenaire-de-la-chaire-industrielle-corenstock/>

<sup>5</sup> <http://chaire-mathematiques-appliquees.emse.fr/>

à créer et déployer des solutions d'Intelligence Artificielle développées pour transformer et améliorer les pratiques industrielles au sein d'un service particulier de l'entreprise : le laboratoire d'analyse des défaillances. Ce laboratoire est un maillon essentiel de la production industrielle, en charge d'analyser et de diagnostiquer les problèmes de qualité et de défaillance qui apparaissent dans les processus de production de wafers micro-électroniques (fabrication de micro-puces) et de produits complexes.

L'analyse des défaillances tout au long de la chaîne de valeur à partir d'une puce, est une condition préalable à l'amélioration de la fiabilité et de la qualité et donc de la compétitivité des dispositifs électroniques, en particulier sur le marché de l'automobile et de l'industrie pour les applications exigeantes en matière de fiabilité et de sécurité. La forte progression de la numérisation et de l'automatisation offre un formidable potentiel pour mieux maîtriser la fiabilité des produits de haute technologie qui sont basés sur des systèmes électroniques de plus en plus complexes.

Grâce à l'apport du digital, l'objectif du projet est donc de transformer les pratiques actuelles du laboratoire d'analyse de défaillance, en mobilisant les techniques du machine learning et deep learning. Ces techniques permettent d'enrichir la capacité explicative des modèles statistiques, par une capacité renforcée à discriminer les causes possibles de non fiabilité et à prendre en compte leurs interactions. Plus généralement, le projet s'inscrit dans les travaux sur l'efficacité du diagnostic par apprentissage, en particulier par classification.

### Problématique, Objectifs et Mission de recherche

La spécificité d'un laboratoire d'analyse des défaillances est d'opérer un certain nombre de tâches de diagnostic électriques, physiques et chimiques, sur les supports micro-électroniques qui contiennent les composants électroniques susceptibles d'être atteint par un défaut (défaillance). Lors de cette analyse, la succession des tâches de diagnostic dépend des résultats observés jusqu'à l'étape  $t$  ; la décision pour le choix de l'étape de diagnostic suivant  $t+1$  dépend des résultats obtenus avant la phase  $t$ , et de l'expertise humaine. Ce processus complet de diagnostic, appliqué à un support électronique sur lequel un défaut a été observé, peut être vu comme un parcours sur un graphe de décision où chaque nœud est un choix parmi un ensemble d'alternatives possibles (alternatives : quelle est l'étape de diagnostic suivante à réaliser ?). Les résultats de chaque tâche de diagnostic se présentent sous la forme de données variées (notamment d'importantes bases de données d'imagerie par microscopies diverses, de signaux physiques, de mesures, de données textuelles ...) qui sont utilisées par l'analyste pour la construction de son diagnostic final : une hypothèse explicative de la défaillance analysée.

En collaboration avec un doctorant et avec l'équipe de recherche, le post-doctorant sera en situation d'assurer le pilotage opérationnel du déroulement de certaines des tâches scientifiques à réaliser durant le projet. Le post-doctorant prendra lui-même en charge les travaux relevant de l'apprentissage supervisés. Les travaux scientifiques à développer viseront à la fois des avancées scientifiques et théoriques et la réalisation de preuves de concepts démontrées à partir de données réalistes issues du milieu industriel, dans le but de favoriser l'intégration ultérieure des résultats par les industriels.

Le post-doctorant travaillera notamment sur les tâches suivantes, qui s'appuieront sur la mise à disposition de données industrielles historiques résultant de la sauvegarde de processus complets d'analyse de défaillance (depuis la requête de diagnostic, jusqu'au diagnostic final) :

- Appliquer des techniques NLP (prétraitement, vectorisation, etc.) et de réduction de dimension (e.g. Variational Auto-Encoder (VAE), Multiple Correspondence Analysis (MCA), etc.) dans le but d'optimiser la représentation de l'espace de décision formé essentiellement de données textuelles et catégorielles.
- Mobiliser les techniques de clustering, de classification et de deep learning dans le but de déterminer des liens de corrélations entre d'une part (i) l'ensemble des données représentant la demande initiale de

diagnostic et d'autre part (ii) l'ensemble des données représentant le diagnostic final résultant des analyses. Ces travaux supposent de traiter des données industrielles pour la plupart catégorielles ou textuelles, et d'analyser comparativement les performances de plusieurs techniques possibles pour l'identification de ces corrélations.

- Articuler puis intégrer les résultats précédents, avec un deuxième volet de travaux en cours de développement par un doctorant de l'équipe de recherche, qui mobilise cette fois-ci non plus les données liées au diagnostic final, mais les données liées aux résultats de chaque phase intermédiaire de diagnostic.

### Profil du candidat

Le candidat en post-doctorat devra disposer des compétences initiales de niveau ingénieur/M2 dans le domaine des mathématiques appliquées, avec une orientation vers les sciences des données, les modèles probabilistes, le machine learning, le deep learning. Il devra avoir soutenu sa thèse au moment du démarrage de poste (les candidats 'ingénieur de recherche', non titulaires de doctorat mais avec une forte expérience dans le domaine peuvent nous contacter directement).

Un parcours de recherche et des expériences dans un ou plusieurs des domaines suivants seront appréciées :

- Mathématiques appliquées, orientées vers l'analyse et le traitement des données ;
- Science des données, Machine learning, Deep learning, identification de modèles.
- Capacités à traiter des données qualitatives et/ou textuelles
- Maîtriser les langages de programmation : Python et R.
- Apprentissage supervisé, Clustering, Classification
- Motivation pour l'application des méthodes développées, notamment pour le diagnostic en contexte industriel.

Bien sûr un intérêt ou des expériences complémentaires en lien avec l'industrie de la microélectronique seront également appréciées. Une capacité à publier en revue scientifique, ainsi qu'un bon niveau d'anglais sont attendus.

### Procédure de candidature

Les dossiers de candidature devront comprendre :

- Un curriculum vitae faisant état des activités d'enseignement, des travaux de recherche et, éventuellement, des relations avec le monde économique et industriel,
- Lettre de Motivation
- A la discrétion des candidats, des lettres de recommandation,
- La copie du doctorat (ou PhD),
- La copie d'une pièce d'identité

Les dossiers de candidature sont à déposer sur la plateforme RECRUITEE le **30 novembre 2021 au plus tard**

URL de dépôt de candidature : <https://institutminestelecom.recruitee.com/o/postdoctorat-en-sciences-de-donnees-institut-henri-fayol>

#### **Pour en savoir plus**

Pr Xavier BOUCHER , tel(Tel: +33 (0)4 77 42 01 33, [boucher@emse.fr](mailto:boucher@emse.fr)

Pr. Mireille BATTON-HUBERT, Tel: +33 (0)4 77 42 00 93, [mireille.batton-hubert@emse.fr](mailto:mireille.batton-hubert@emse.fr)

Pour tout renseignement administratif, s'adresser à :

Amandine HIRONDEAU

Tel + 33 (0)4 77 42 01 03 , Mel: [hirondeau@emse.fr](mailto:hirondeau@emse.fr)