

Fiche de poste : Post-doctorat en découverte causale et quantification d'incertitudes pour les lignes de lumière

L'environnement

Le DM2S (Département de Modélisation des Systèmes et Structures) développe des outils de simulation pour la conception et l'évaluation de systèmes dans les disciplines de base du nucléaire, i.e. thermohydraulique, thermomécanique et neutronique, toutes filières confondues. Au sein du DM2S, et du SGLS (Service de Génie Logiciel pour la Simulation), le LIAD (Laboratoire d'Intelligence Artificielle et de science des Données) réalise et maintient une plateforme générique, pérenne et open source pour fournir aux physiciens des méthodes et outils leur permettant d'améliorer leurs modèles, d'optimiser leurs conceptions et de traiter les incertitudes de leurs études : la plateforme Uranie. Cette plateforme permet, dans l'approche VVQI (Validation, Vérification et Quantification d'Incertitude), de créer des plans d'expériences adaptés aux besoins d'une analyse de sensibilité, d'un problème d'optimisation ou de la génération d'une base d'apprentissage ou de test pour un modèle de substitution.

Le poste

Au sein du laboratoire LIAD (Laboratoire d'Intelligence Artificielle et de science des Données) appartenant au service SGLS (Service de Génie Logiciel pour la Simulation), vous évoluez dans une équipe spécialisée dans le traitement des incertitudes en simulation numérique. Vous contribuez au renforcement de ces activités au travers du post-doctorat DALLIAE dont l'objectif est la Détection d'Anomalies sur les Lignes de Lumière du synchrotron par Intelligence Artificielle Explicable.

Le post-doctorat DALLIAE vise à proposer une méthode générique basée sur les graphes causaux (bayésiens) [1] afin de détecter les anomalies lors des expériences sur les lignes de lumières et leur interprétabilité. Parmi les graphes causaux, on s'intéressera en particulier aux graphes dirigés acycliques (DAG) [2]. L'enjeu est de trouver les liens causaux simples (un paramètre) et joints (combinaison de paramètres) les plus pertinents pour caractériser les causes d'une anomalie. On s'intéressera également à la quantification des incertitudes associées aux liens causaux identifiés afin de s'assurer de leur pertinence. Cette recherche de causalité est d'autant plus difficile du fait de la variété des instruments, des paramètres [2], de leur modification au cours de l'expérience [3], du nombre combinatoire des effets joints à étudier et de la sous-représentation des anomalies dans les données.

En pratique, cette méthode permettra de limiter l'impact des anomalies de fonctionnement des grands instruments de rayon X ou laser dont il est nécessaire de connaître les liens entre les caractéristiques du faisceau et les paramètres physiques de l'optique des lignes. Des anomalies/variations soudaines ou lentes peuvent ainsi être observées avec le temps comme des aberrations de focalisation qui affectent directement la qualité et la rapidité des mesures. Ainsi, pouvoir comprendre et caractériser les causes de ces dysfonctionnements et des écarts au fonctionnement optimal de la chaîne de mesures est un enjeu majeur pour rétroagir rapidement et garantir une fiabilité maximale d'exploitation de la ligne de lumières ou de laser. Bien qu'il existe dans la littérature en intelligence artificielle de nombreuses méthodes de détection d'anomalies, elles sont généralement basées sur la corrélation qui est peu efficace pour traduire les relations de cause à effet.

Ce post doctorat s'inscrit dans le cadre d'une collaboration entre plusieurs équipes du CEA qui amènent

chacune des compétences distinctes telles que l'IA, la physique liée aux lignes de lumière et l'instrumentation dont le NRX (Équipe Nanostructures et Rayons X à Grenoble) et le LMJ (Laser Mégajoule) à Bordeaux.

L'objectif du post-doctorat est ainsi de proposer une IA interprétable (les graphes causaux), qui sera un soutien aux opérateurs et scientifiques des lignes de lumières. Elle assistera la résolution de pannes et complétera l'arsenal de diagnostic pour une correction ou rétroaction dans des temps adaptés.

Il s'agit ainsi de développer un modèle basé sur la causalité pour déterminer les paramètres des capteurs impliquant les anomalies. Ainsi, votre travail de recherche s'articulera autour de quatre axes:

- Comprendre et prendre en main les données produites par les lignes de lumière : les paramètres des instruments de mesure et les différents types d'anomalies. Pour ce faire, on s'appuiera sur les connaissances des partenaires du projet, experts en physique, en optique et en instrumentation liées aux lignes de lumières.
- Concevoir le modèle basé sur les graphes causaux explicitant les liens entre les différents paramètres et les anomalies. On s'intéressera en particulier à la détection des variables latentes impactant l'incertitude liée à la détection d'un lien causal, la dimensionnalité des données et aux interactions entre les différentes potentielles causes. Pour ce faire, on pourra se placer dans le cadre des graphes bayésiens causaux.
- Soutenir et participer au travail du LIAD en prenant part aux sollicitations qu'elles soient côté méthodologie incertitude, intelligence artificielle (voire les deux) ;
- Valoriser vos travaux via la rédaction de notes techniques, l'écriture de publications dans des conférences et des journaux consacrés, et la participation aux réunions du projet DALLIAE avec l'ensemble des experts.

Durée : 24 mois

Début du projet : 4^{ème} trimestre 2023-1^{er} trimestre 2024

Profil du Candidat

De profil docteur/e en IA (statistique), une expérience en causalité ou quantification d'incertitudes est appréciée mais pas obligatoire.

La capacité à travailler en équipe est nécessaire, tout en faisant preuve d'autonomie dans les tâches au quotidien. Les développements en IA avançant rapidement, il est nécessaire que la candidate ou le candidat aient la capacité de renouveler et enrichir ses compétences de manière continue. Enfin, d'excellentes capacités d'expression, orale et écrite, en anglais, vous seront nécessaires pour échanger avec les nombreux partenaires du projet.

Sur le plan scientifique et technique, vous pouvez justifier :

- De solides compétences en *machine learning* (graphe bayésien et/ou quantification d'incertitudes) et en statistique,
- D'une expérience significative en développement logiciel, s'appuyant idéalement sur
 - la maîtrise d'un langage de prototypage (idéalement Python 3)
 - l'utilisation d'outils de développement et de partage de code tels que Git ou SVN sera également appréciée.

Contact : aurore.lomet@cea.fr

Bibliographie

- [1] J. Peters, D. Janzing et B. Schölkopf, Elements of causal inference: foundations and learning algorithms, The MIT Press, 2017.
- [2] C. Glymour, K. Zhang et P. Spirtes, «Review of causal discovery methods based on graphical models,» Frontiers in genetics, vol. 10, p. 524, 2019.
- [3] B. Huang, K. Zhang, J. Zhang, J. Ramsey, R. Sanchez-Romero, C. Glymour et B. Schölkopf, «Causal discovery from heterogeneous/nonstationary data,» The Journal of Machine Learning Research, vol. 21, p. 3482–3534, 2020.
- [4] A. Arzac, A. Lomet et J.-P. Poli, «Causal discovery for time series with constraint-based model and PMIME measure,» arXiv preprint arXiv:2305.19695, 2023.