Contenus

Code du Cours	Nom du Cours	Semestre du Cours	Cours Théoriques	Travaux Dirigés (TD)	Travaux Pratiques (TP)	Crédit du Cours	ECTS
INF 539	Intelligence artificielle explicable	2	3	0	0	3	6

Cours Pré-Requis	
Conditions d'Admission au Cours	

Langue du Cours	Anglais
Type de Cours	Électif
Niveau du Cours	Master
Objectif du Cours	Ce cours porte sur l'explication et l'interprétation des décisions des algorithmes d'apprentissage automatique. Il vise principalement à initier les étudiants aux méthodes d'intelligence artificielle explicable (XAI) et à montrer, à travers des applications pratiques, comment ces méthodes sont utilisées dans divers domaines.
Contenus	Ce cours vise à interpréter les décisions, les prédictions et les inférences des systèmes d'intelligence artificielle, ainsi qu'à expliquer comment et pourquoi ces résultats sont produits par les algorithmes existants. Il offre une vue d'ensemble de l'interprétation des décisions des modèles d'apprentissage automatique, utilisés dans divers domaines allant de la santé à la finance, souvent qualifiés de « boîtes noires », et aborde les aspects essentiels du développement de systèmes d'IA fiables, transparents et éthiques. Les étudiants auront l'occasion d'appliquer les méthodes présentées dans le cours à l'aide de Python et de discuter leurs résultats.
Ressources	- Mehta, M., Palade, V., & Chatterjee, I. (Eds.). (2023). Explainable AI: Foundations, methodologies and applications (Vol. 232, p. 273). Springer. - Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (Eds.). (2019). Explainable AI: interpreting, explaining and visualizing deep learning (Vol. 11700). Springer Nature. - Molnar, C. (2020). Interpretable machine learning. - Hsieh, W., Bi, Z., Jiang, C., Liu, J., Peng, B., Zhang, S., ... & Liu, M. (2024). A comprehensive guide to explainable AI: from classical models to LLMs. arXiv preprint arXiv:2412.00800.

Intitulés des Sujets Théoriques

Semaine	Intitulés des Sujets
1	Concepts Fondamentaux : Explicabilité, Transparence, Interprétabilité, Équité, Robustesse et IA explicable
2	Fondements Théoriques de l'IA Explicable
3	Interprétabilité des Modèles d'Apprentissage Automatique Traditionnels
4	Interprétabilité des Modèles d'Apprentissage Profond
5	Techniques pour IA explicable
6	Méthodes d'Attribution des Caractéristiques
7	Techniques pour la Visualisation
8	Examen Partiel
9	Techniques de Traitement des Données Temporelles et Séquentielles
10	Explicabilité Multimodale
11	Applications de l'IA Explicable - Partie I
12	Applications de l'IA Explicable - Partie II
13	Défis Rencontrés
14	Présentations des Etudiants