Université du Québec en Outaouais **Département d'informatique et d'ingénierie**

Sigle : CYB1183 Gr. 20
Titre : Sécurité et intelligence artificielle
Session : Automne 2026 Horaire et local
Professeur.e : Bouhaddi, Myria

1. Description du cours paraissant à l'annuaire :

Objectifs

Au terme de ce cours, l'étudiant.e aura acquis des compétences et des capacités dans la résolution de problèmes de sécurité des algorithmes de l'intelligence artificielle et son application en cybersécurité, autant au niveau des cyberattaques utilisant l'intelligence artificielle que des approches basées sur l'apprentissage automatique pour produire des systèmes de cybersécurité autonomes et intelligents.

Contenu

Vulnérabilités et risques spécifiques des algorithmes de l'Intelligence Artificielle (I.A.). Taxonomie des attaques des systèmes informatiques basés sur l'I.A. (ex.: empoisonnement des données, inférence des données d'apprentissage, inférence des paramètres des modèles). Méthodes de sécurisation des algorithmes de l'I.A. Cyberattaques basées sur l'I.A. : générer des attaques personnalisées, analyse automatique de données post-infiltration, outils open source pour développer des algorithmes de l'I.A. malveillants, etc. Cyberdéfense basée sur l'I.A. : classification de données pour la détection d'anomalies, cyberdéfense autonome et adaptative, outils de défense basés sur l'IA, etc. Ce cours comporte des séances obligatoires de travaux dirigés (TD).

Descriptif – Annuaire

2. Objectifs spécifiques du cours :

Au terme de ce cours, l'étudiant·e sera en mesure de :

1. **Comprendre les principaux concepts et méthodes d'intelligence artificielle** applicables à la cybersécurité.
2. **Appliquer des techniques d'apprentissage automatique et profond** à des problèmes de cybersécurité tels que la détection d'intrusions, d'anomalies et de comportements malveillants.
3. **Concevoir et évaluer des mécanismes de détection, de décision et de réponse** basés sur l'intelligence artificielle.
4. **Identifier les vulnérabilités et les attaques visant les systèmes d'IA et appliquer des mécanismes de défense adaptés.**

3. Stratégies pédagogiques :

La structure pédagogique de ce cours intègre les éléments suivants :

1. **Cours magistraux** : une séance hebdomadaire de 3 heures.
2. **Devoirs** : répartis tout au long du semestre pour renforcer les concepts abordés.
3. **Projet avec présentation orale** : permettant aux étudiants de démontrer leur compréhension et d'appliquer les connaissances acquises.

Deux examens : Évaluant la maîtrise des compétences et des connaissances couvertes en cours.

4. Heures de disponibilité ou modalités pour rendez-vous :

Communication par courriel (myria.bouhaddi@uqo.ca) et via le forum de discussion.

5. Plan détaillé du cours sur 15 semaines :

Semaine	Thèmes	Dates
1	**Introduction à l'IA appliquée à la cybersécurité**	8 sept.

	• Présentation de la progression du cours et des TP • Rôle de l'IA dans la cybersécurité moderne • Typologie des problèmes : détection, classification, prédiction et décision • Principales applications de l'IA en cybersécurité	
2	**Fondements de l'apprentissage automatique** • Données, observations, variables (*features*) et labels • Apprentissage supervisé, non supervisé et introduction à l'apprentissage par renforcement • Jeu de données d'entraînement, de validation et de test • Pipeline ML : collecte, préparation, entraînement, évaluation et déploiement **Travail pratique 1 :** Préparation et exploration d'un jeu de données de cybersécurité	15 sept.
3	**Méthodes classiques d'apprentissage automatique en cybersécurité** • Principe de la classification et principales méthodes : régression logistique, arbres de décision, Random Forest et SVM • Choix d'un algorithme selon le problème de cybersécurité • Évaluation : matrice de confusion, accuracy, précision, rappel, F1-score • Importance des faux positifs et faux négatifs en cybersécurité • **Étude de cas :** classification d'intrusion réseau **Devoir 1** **Travail pratique 2 :** classification d'un trafic réseau	22 sept.
4	**Détection d'intrusions et détection d'anomalies (partie 1)** • Classification vs détection d'anomalies • Détection par signatures vs détection comportementale • Comportement normal, anomalie et données atypiques • Introduction aux méthodes non supervisées et au clusterings	29 sept.
5	**Détection d'intrusions et détection d'anomalies (partie 2)** • K-means et Isolation Forest • Détection d'anomalies appliquée au trafic réseau • Limites : faux positifs, nouvelles attaques et évolution du trafic • Introduction à la dérive (*concept drift*) • **Étude de cas :** Détection de comportements inconnus **Travail pratique 3 :** Détection d'anomalies réseau	6 oct.
6	**SEMAINE D'ÉTUDES**	13 oct.
7	**EXAMEN DE MI-SESSION**	20 oct.
8	**Introduction aux réseaux de neurones** • Neurone artificiel : entrées, poids, biais et fonction d'activation • Architecture d'un réseau de neurones : couches d'entrée, cachées et de sortie	27 oct.

	• Principe de l'entraînement, fonction de perte et surapprentissage • Introduction au Deep Learning et applications en cybersécurité • **Étude de cas :** détection/classification de malwares **Travail pratique 4 :** Détection/classification de malwares par réseau de neurones	
9	**Apprentissage profond appliqué à la cybersécurité** • Du ML classique au Deep Learning • Réseaux neuronaux profonds et CNN • RNN/LSTM pour les données séquentielles • Choix d'une architecture selon la nature des données? • **Étude de cas :** analyse de séquences et de journaux de sécurité **Devoir 2** **Travail pratique 5 :** Analyse de séquences ou de logs avec un modèle profond	3 nov.
10	**Autoencodeurs et détection avancée d'anomalies** • Principe d'un autoencodeur • Encodeur, représentation latente et décodeur • Reconstruction et erreur de reconstruction • Détermination d'un seuil pour la détection d'anomalies • **Étude de cas :** IDS basé sur l'erreur de reconstruction **Travail pratique 6 :** Autoencodeur pour la détection d'anomalies	10 nov.
11	**Introduction à l'apprentissage par renforcement en cybersécurité** • De la classification à la prise de décision séquentielle • Agent, environnement, états, actions et récompenses • Politique de décision et interaction agent-environnement • Exploration/exploitation et limites de la décision autonome • **Étude de cas : gestion adaptative d'une menace**autonome **Travail pratique 7 :** Agent de réponse en cybersécurité	17 nov.
12	**Systèmes de défense et de réponse automatisés basés sur l'IA** • De la détection à la réponse • Défenses réactives et adaptatives et introduction aux systèmes SOAR • Confiance dans la décision et coût des erreurs de classification • Intervention humaine et *human-in-the-loop* • **Étude de cas :** scénarios de réponse automatisée à une menace	24 nov.
13	**Vulnérabilités, attaques et défenses des systèmes d'IA** • Vulnérabilités et surface d'attaque des systèmes d'apprentissage automatique • Attaques : empoisonnement, évasion, *Membership Inference* et extraction de modèle • Défenses : mécanismes de robustesse et confidentialité différentielle • Impacts sur la confidentialité, l'intégrité et la disponibilité • **Étude de cas :** attaque *Membership Inference* basée sur des *Shadow Models* et défense par confidentialité différentielle	1 déc.

	Travail pratique 8 – Attaque et défense d'un modèle d'apprentissage automatique	
14	Présentation des projets	8 déc.
15	Examen final	15 déc.

6. Évaluation du cours :

L'évaluation est l'appréciation du niveau d'apprentissage atteint par l'étudiant(e) par rapport aux objectifs des cours et des programmes.

Dans le cas spécifique du cours **sécurité et intelligence artificielle,** l'attribution des notes se fera selon la répartition suivante :

- **Examen de mi-session : 25 %**
- **Examen final : 25 %**
- **Travail de session : 30 %**
- **Devoir 1 et 2 : 20 %**

7. Politiques départementales et institutionnelles :

- Politique du département d'informatique et d'ingénierie relative à la tenue des examens
- Note sur le plagiat et sur la fraude
- Politique relative à la qualité de l'expression française écrite chez les étudiants et les étudiantes de premier cycle à l'UQO
- Absence aux examens : cadre de gestion, demande de reprise d'examen (formulaire)

Tolérance ZÉRO en matière de violence à caractère sexuel.

Le Bureau d'intervention et de prévention en matière de harcèlement (BIPH) a pour mission d'accueillir, soutenir et guider toute personne vivant une situation de harcèlement, de discrimination ou de violence à caractère sexuel. Le BIPH oriente ses actions afin de prévenir les violences à caractère sexuel pour que nous puissions étudier, travailler et s'épanouir dans un milieu sain et sécuritaire.

Vous vivez ou êtes une personne témoin d'une situation de violence à caractère sexuel ? Vous êtes une personne membre de la communauté étudiante ou une personne membre du personnel, autant à Gatineau qu'à Ripon et St-Jérôme, l'équipe du BIPH est là pour vous, sans jugement et en toute confidentialité.

Ensemble, participons à une culture de respect.

Pour de plus amples renseignements consultez UQO.ca/biph ou écrivez-nous au Biph@uqo.ca

8. Principales références :

1. Diogenes, Yuri, and Erdal Ozkaya. *Cybersecurity-attack and defense strategies: Infrastructure security with red team and blue team tactics*. Packt Publishing Ltd, 2018.
2. Chio, Clarence, and David Freeman. Machine learning and security: Protecting systems with data and algorithms. " O'Reilly Media, Inc.", 2018.
3. Halder, Soma, and Sinan Ozdemir. Hands-On Machine Learning for Cybersecurity: Safeguard your system by making your machines intelligent using the Python ecosystem. Packt Publishing Ltd, 2018.
4. Alessandro Parisi. Hands-On Artificial Intelligence for Cybersecurity: Implement smart AI systems for preventing cyber attacks and detecting threats and network anomalies. Packt Publishing Ltd, 2019.

9. Page Web du cours :

https://moodle.uqo.ca