

# ASC2-2024-IA2P-1- Génération de données synthétiques en infra-rouge

Niveau requis	Durée	Mots-clés
bac+5	6 mois	modélisation 3D, simulation, infra-rouge, détection de cibles, IA

Description du stage
<p>Le stagiaire sera chargé de concevoir et mettre en place des outils de simulation d'images infrarouges en utilisant des logiciels open-source tels que Blender ou Unreal Engine. L'objectif principal de ce projet consiste à enrichir les bases de données de simulation existantes, en générant un grand volume de données visuellement fidèle en infrarouge.</p>

Compétences indispensables	Compétences souhaitées
Rendu 3D	connaissance sommaire du machine learning

Observations / Les "+" du stage

# ASC2-2024-IA2P-2- Génération automatique de texte pour un domaine cible par l'utilisation conjointe de grands modèles de langues (LLMs) et de techniques de récupération d'informations (RI).

Niveau requis	Durée	Mots-clés
bac+5	6 mois	Grands modèles de langue, LLMs, base de données vectorielle, base de données sémantique, prompt-engineering, récupération d'information, RI, RAG, IA, Transformers

Description du stage
<p>Les grands modèles de langues, communément appelés LLMs, peuvent être utilisés pour générer automatiquement du texte en lien avec un sujet ou une actualité spécifique. Les techniques de prompting permettent de faire varier le contenu généré en influant, par exemple, sur le registre de langue (familier/courant/soutenu). Afin de gagner en fiabilité, le modèle de langue peut s'appuyer sur des sources externes provenant d'un domaine d'intérêt pour étoffer les réponses. Les méthodes de récupération d'informations, avec par exemple RAG (Retrieval Augmented Generation), permettent de requêter des informations contenues dans une base de données vectorielle pour consolider le prompt et ainsi produire des réponses plus développées par le LLM. Cela permet de pouvoir utiliser des données à jour sur un domaine cible sans réentraînement et limite également les potentielles hallucinations. L'objectif du stage est de mettre en place ce type d'architecture en s'appuyant les nombreux outils open-source. Plusieurs étapes sont nécessaires :</p> <p>Etat de l'art sur les méthodes de récupération d'informations.</p> <p>Génération de la base de données sémantique via une méthode identifiée dans l'état de l'art.</p> <p>Vérification de la cohérence sémantique de la base.</p> <p>Génération de réponses augmentées via l'utilisation conjointe des informations requêtées et d'un LLM génératif.</p> <p>Analyse des résultats, comparaison avec/sans apport de sources externes.</p>

Compétences indispensables	Compétences souhaitées
Deep Learning, Transformers, Python	Connaissances en traitement automatique du langage

Observations / Les "+" du stage
Utilisation de technologies à l'état de l'art pour la génération de texte

# ASC2-2024-IA3D-3- Recherche de similarité entre binaires par intelligence artificielle

Niveau requis	Durée	Mots-clés
bac+5	6 mois	Similarité de binaires, Large Language Models, Graph Neural Networks, Reverse Engineering

Description du stage
<p>La recherche de similarité entre binaires est un sujet complexe au vu du nombre de paramètres différents à prendre en compte (compilateur utilisé, architecture, niveau d'optimisation, version du compilateur, etc.). Ce problème est particulièrement intéressant dans le domaine du reverse engineering car il permet de reconnaître des fonctions connues (et donc fait gagner du temps aux reverseurs) mais également de faire de l'attribution de code ou encore de la recherche de vulnérabilités connues.</p> <p>Pour cela, de nombreuses méthodes à base d'intelligence artificielle ont vu le jour dans le domaine de la recherche. Ces méthodes sont principalement basées sur des graphes ou du traitement du langage [1][2].</p> <p>Une étude de ces méthodes a déjà été réalisée par le département IA3D mettant en avant l'intérêt de l'utilisation de méthodes à base d'intelligence artificielle. L'objectif du stagiaire sera donc de participer à la suite de cette étude qui portera sur :</p> <ul style="list-style-type: none"><li>- L'entraînement d'un modèle de langue sur un grand nombre de données assembleur</li><li>- La création d'un dataset pertinent et diversifié d'un grand nombre de binaires</li><li>- L'étude et l'implémentation de la fusion entre un modèle de langue et un modèle de graphes</li></ul> <p>Le stage sera encadré par le laboratoire Datascience du département Intelligence Artificielle pour les Agents Autonomes &amp; Datascience (IA3D).</p> <p>[1] Marcelli, Andrea, et al. "How machine learning is solving the binary function similarity problem." 31st USENIX Security Symposium (USENIX Security 22). 2022.</p> <p>[2] <a href="https://github.com/SystemSecurityStorm/Awesome-Binary-Similarity">https://github.com/SystemSecurityStorm/Awesome-Binary-Similarity</a></p>

Compétences indispensables	Compétences souhaitées
Programmation Python, Natural Language Processing	Théorie des graphes, connaissances en reverse engineering et langage assembleur (facultatif)

Observations / Les "+" du stage
Travail sur un sujet de recherche en plein essor. Utilisation de technologies à l'état de l'art (transformers et graph neural networks)

# ASC2-2024-IA3D-4- Apprentissage auto-supervisé pour les réseaux de neurones de graphes

Niveau requis	Durée	Mots-clés
Bac +5	6 mois	Graph Neural Network, Self Supervised Learning, Deep Learning

Description du stage
<p>Le Machine Learning appliqué aux graphes est devenu un sujet majeur de l'intelligence artificielle et a permis de créer des solutions réelles activement déployées à grande échelle. Les graphes fournissent un cadre général et flexible pour décrire et analyser n'importe quel ensemble d'entités et leurs interactions. De plus, la capacité des GNN (Graph Neural Networks) à exprimer les relations entre les données en font un outil de prédilection pour de nombreuses applications (Systèmes de recommandations, Graphs de connaissances, Biologie, Cyber-sécurité, ...).</p> <p>Cependant, les travaux passés ont largement été concentrés sur l'apprentissage supervisé, fortement dépendant de la labélisation des données, peu généralisable et peu robuste. C'est pourquoi, inspiré par son succès dans le traitement du langage naturel et la vision par ordinateur, le Self Supervised Learning est devenu l'approche la plus prometteuse pour l'entraînement de GNN, mais se heurte à des difficultés d'adaptation inhérentes à l'apprentissage sur graphes.</p> <p>Le département IA3D (Intelligence Artificielle pour les Agents Autonomes &amp; Datascience) propose un stage sur les méthodes de Self Supervised Learning appliquées aux données sous forme de graphes. Les objectifs du stage sont multiples et permettront au stagiaire de monter en compétence en méthodologie scientifique et en Machine Learning :</p> <ol style="list-style-type: none"><li>1. Établir un état de l'art des méthodes existantes et évaluer les performances de ces méthodes sur des données publiques pour comprendre leurs points forts et leurs limites ;</li><li>2. Évaluer ces méthodes sur un jeu de données de programmes informatiques sous forme de graphes ;</li></ol>

Compétences indispensables	Compétences souhaitées
Programmation Python, Deep Learning	Maitrise d'un framework de Deep Learning (Pytorch, Tensorflow)

Observations / Les "+" du stage
Le stagiaire travaillera sur des méthodes à l'état de l'art, et sera accompagné par des experts qui ont travaillé sur le sujet.

# ASC2-2024-IA3D-5- IA pour la navigation en flotte de robots terrestres

Niveau requis	Durée	Mots-clés
Bac+5	6 mois	Robotique, Apprentissage par renforcement, navigation, trajectoire, mutli-agents, comportement en flotte

Description du stage
<p>La navigation autonome d'une flotte de robot est un sujet d'actualité en robotique. Les avancées rapides de ces dernières années ont donné naissance à des algorithmes d'intelligence artificielle très performants. Il existe néanmoins des problématiques propres au ministère des armées qui demandent une adaptation de ces IA (discrétion des communications, nature des terrains...)[1]</p> <p>L'objectif de ce stage est d'évaluer les algorithmes actuellement disponibles. Le support d'évaluation utilisé est la plateforme robotique MuSHR [2] disponible en version physique et en simulation en exploitant le logiciel de simulation Nvidia ISAAC. Des algorithmes de navigation autonome sont déjà disponibles [3][4], le développement doit se focaliser sur l'approche multi-agents.</p> <p>Le stage se déroulera en 2 étapes :</p> <ul style="list-style-type: none"><li>- Exploitation de la simulation pour réaliser des apprentissages de génération et suivi de trajectoire en flotte.</li><li>- La validation du fonctionnement lors d'une expérimentation pour quantifier le « reality gap »</li></ul> <p>L'encadrement sera réalisé par le département Intelligence Artificielle pour les Agents Autonomes &amp; Data-science (IA3D).</p> <p>[1] Bois, J., Puig, A., Rullière, L., Teboul, Y., Ossola, M., Kotenkoff, A., and Formoso, M. (2022). Heterogeneous swarming for collaborative combat using Multi-agent Deep Reinforcement Learning. In Conference on Artificial Intelligence for Defense, Actes de la 4ème Conference on Artificial Intelligence for Defense (CAID 2022), Rennes, France. DGA Maîtrise de l'Information.</p> <p>[2] MuSHR: Multi-agent System for non-Holonomic Racing, <a href="https://github.com/prl-mushr/mushr">https://github.com/prl-mushr/mushr</a></p> <p>[3] Elbanhawi, M. and Simic, M. (2014). Sampling-based robot motion planning : A review. IEEE Access, vol. 2 :56-77.</p> <p>[4] Cimurs, R., Suh, I. H., and Lee, J. H. (2022). Goal-driven autonomous exploration through deep reinforcement learning. IEEE Robotics and Automation Letters, vol. 7 (n°2) :730-737.</p>

Compétences indispensables	Compétences souhaitées
Programmation Python, machine learning	Apprentissage par renforcement, robotique

Observations / Les "+" du stage
Présence d'une maquette réelle. Expérimentation en plus de la modélisation sur ordinateur

# ASC2-2024-IA2P-6- Détection de cibles sur images SAR

Niveau requis	Durée	Mots-clés
bac+5	6 mois	détection, SAR, image, Deep Learning

Description du stage
<p>Les images radar (SAR) ont l'avantage de pouvoir être obtenues le jour et la nuit ainsi que dans des conditions météo variées. Des satellites permettent d'acquérir de telles images, en mer ou sur terre. Ces images sont de plus mauvaise qualité que les images optiques et les annotations ne sont pas toujours fiables. Nous pensons orienter ce stage sur l'étude du dataset XViewSAR3, dataset d'imagerie SAR satellite maritime contenant des bateaux.</p> <p>Ce stage visera à établir un benchmark de détection sur ce dataset ainsi que d'explorer certaines pistes permettant de prendre en compte la fiabilité des annotations données.</p> <p>Les outils pressentis pour ce stage sont Keras/Tensorflow ou Pytorch, la connaissance du langage python est donc souhaitée.</p>

Compétences indispensables	Compétences souhaitées
Connaissances en deep learning	

Observations / Les "+" du stage

# ASC2-2024-IA3D-7- Génération procédurale d'environnement d'entrainement

Niveau requis	Durée	Mots-clés
Bac+5	6 mois	Génération procédurale, renforcement learning, RL, Deep Learning,

Description du stage
<p>L'apprentissage par renforcement nécessite une multitude d'environnements d'entrainement pour espérer généraliser les choix effectués par un agent pendant son apprentissage. Il s'agit d'évaluer la pertinence d'algorithmes de type Wave function Collapse pour générer une diversité d'environnements à partir de quelques exemples.</p> <p>Exemple d'utilisation de l'algorithme: <a href="https://www.youtube.com/watch?v=zIRTOgfsjl0">https://www.youtube.com/watch?v=zIRTOgfsjl0</a> <a href="https://www.youtube.com/watch?v=2SuvO4Gi7uY">https://www.youtube.com/watch?v=2SuvO4Gi7uY</a></p> <p>Le stagiaire devra</p> <ol style="list-style-type: none"><li>1) Définir les critères d'évaluation de la qualité des environnement généré.</li><li>2) Trouver un/des générateurs d'environnement comme base de comparaison</li><li>3) Implémenter l'algorithme pour les environnements d'évaluation</li><li>4) Evaluer les environnements générés et dégager les avantages et limites de cette approche.</li></ol>

Compétences indispensables	Compétences souhaitées
	python

Observations / Les "+" du stage

# ASC2-2024-IA2P-8- Débruitage d'images SAR et estimation de paramètres capteur

Niveau requis	Durée	Mots-clés
Bac+5	6 mois	Débruitage, image, SAR, deep learning, régression paramètres

## Description du stage

Les images radar (SAR) sont utilisées parce qu'elles peuvent être obtenues le jour et la nuit et dans beaucoup de conditions climatiques. Ces images ont des bruit caractéristiques issues du capteur utilisé. Afin d'être capable de débruiter ces images, un jeu de données simulées a été généré. De nombreux paramètres capteurs ont été utilisés pour la génération de ces images simulées et elles sont associées à des images débruitées. Le candidat devra dans un premier temps créer un modèle capable de débruiter les images (apprentissage supervisé). Dans un second temps, il pourra mettre à profit les métadonnées pour estimer les paramètres capteurs des images bruitées. Une application du débruitage d'images pourra être la comparaison de résultats de classification.

Compétences indispensables	Compétences souhaitées
Deep learning, python	traitement d'image, théorie du signal

## Observations / Les "+" du stage

Jeu de données conséquent et riche en métadonnées. Liberté d'explorer plutôt l'estimation de paramètres capteurs ou l'amélioration de résultats de classification.



# ASC2-2024-IA2P-9- Backdoor IA

Niveau requis	Durée	Mots-clés
Bac+5	6 mois	deep learning, backdoor, sécurité

Description du stage
<p>Les gains de performance des modèles à l'état de l'art s'accompagnent souvent d'une inflation des architectures/nombres de paramètres. L'entraînement devient alors prohibitif et l'alternative de fine-tuner un modèle pré-entraîné devient pertinente.</p> <p>Or, les frontières de décision des modèles de ML/DL étant pilotées par les couples {donnée, annotation}, il est tout à fait possible d'imposer un comportement au modèle en fonction d'un trigger choisi par l'attaquant (ex : Hidden Trigger Backdoor Attacks. <a href="https://arxiv.org/pdf/1910.00033.pdf">https://arxiv.org/pdf/1910.00033.pdf</a>).</p> <p>Ce stage propose d'explorer la thématique des backdoors IA, en particulier sous le prisme des interactions avec les autres thèmes de la sécurité des IA :</p> <ul style="list-style-type: none"><li>- Certification (Availability Attacks Against Neural Network Certifiers Based on Backdoors, <a href="https://arxiv.org/pdf/2108.11299.pdf">https://arxiv.org/pdf/2108.11299.pdf</a>)</li><li>- Attaques adversariales (Gotta Catch 'Em All: Using Honeypots to Catch Adversarial Attacks on Neural Networks, <a href="https://arxiv.org/pdf/1904.08554.pdf">https://arxiv.org/pdf/1904.08554.pdf</a>)</li><li>- Attaque d'inférence (Membership Inference via Backdooring, <a href="https://arxiv.org/pdf/2206.04823.pdf">https://arxiv.org/pdf/2206.04823.pdf</a>)</li><li>- etc.</li></ul> <p>Le plan proposé :</p> <p>En fonction des sorties de l'état de l'art, un modèle de menace sera alors fixé pour délimiter l'étude (proportion du jeu de données à attaquer, accès aux données et/ou à l'entraînement, etc.), une stratégie pour chaque étape {attaque, détection, défense} sera identifiée et mise en place.</p>

Compétences indispensables	Compétences souhaitées
Deep learning, python	cybersécurité

Observations / Les "+" du stage
<p>Le stagiaire intégrera une équipe technique spécialisée en IA et couvrant un large périmètre : traitement des langues, image, signal, robotique, cyber, etc.</p> <p>Le sujet est exploratoire et permet une grande liberté d'exploration.</p>

# ASC2-2024-IA2P-10- Méthodes Deep Learning pour la modélisation des fonds marins à partir de données issues de navires hydrographiques

Niveau requis	Durée	Mots-clés
Bac+5	6 mois	deep learning, détection d'anomalies, nuages de points

Description du stage
<p>La maîtrise et la protection des fonds marins constituent des enjeux majeurs pour la France, deuxième zone économique mondiale.</p> <p>Afin de cartographier précisément les fonds marins, le Service Hydro-Océanographique de la Marine est appelé à traiter un volume de donnée toujours plus important.</p> <p>Le traitement de ces données requiert actuellement un lourd traitement manuel par les hydrographes. Ainsi, nous souhaiterions explorer la capacité des méthodes de Deep Learning à automatiser certaines de ces tâches.</p> <p>Le contexte du stage est la modélisation de la surface sous-marine à partir des données issues des navires hydrographiques.</p> <p>Plus précisément, il s'agira de détecter les données aberrantes à l'aide de réseaux de neurones. Des travaux ayant déjà initiés en ce sens, le stage visera notamment à explorer de nouvelles représentations des données d'entrée. Il s'en suivra une comparaison avec les méthodes déjà développées au sein de l'équipe.</p> <p>Déroulement du stage:</p> <ul style="list-style-type: none"><li>- Familiarisation avec les problématiques métier</li><li>- Prise en main des données (nuages de points)</li><li>- Formation aux modèles déjà développés au sein de l'équipe (détection d'anomalies avec des U-Net)</li><li>- Exploration d'une représentation alternative des données, et développement de nouveaux modèles</li><li>- Analyse des résultats au regard des problématiques métier</li><li>- Comparaison des surfaces obtenues avec les surfaces de référence</li></ul>

Compétences indispensables	Compétences souhaitées
deep learning	

Observations / Les "+" du stage
<p>Le stagiaire intégrera une équipe technique spécialisée en IA et couvrant un large périmètre : traitement des langues, image, signal, robotique, cyber, etc.</p> <p>Le sujet est exploratoire et permet une grande liberté d'exploration.</p>

# ASC2-2024-IA3D-11- Évaluation des risques de collision de satellites à base de réseaux bayésiens

Niveau requis	Durée	Mots-clés
Bac+5	6 mois	deep learning, bayésien, spatial

Description du stage
<p>Le domaine spatial est un enjeu majeur pour le ministère des armées. Les trajectoires de millions d'objets spatiaux, principalement des débris, sont suivies pour anticiper des éventuelles collisions avec des satellites. Des manœuvres d'évitement sont alors lancées, ce qui a un coût sur le carburant des satellites, donc sur leur durée de vie.</p> <p>L'ESA a lancé un challenge en 2019 portant sur la détection de ces événements rares, à partir des trajectoires observées, à partir d'un dataset publié pour l'occasion.</p> <p>Le département IA3D (Intelligence Artificielle pour les Agents Autonomes &amp; Datascience) propose un stage sur une méthode base sur les réseaux bayésiens pour prédire la probabilité de collision au fur et à mesure des observations, ainsi qu'une incertitude associée.</p> <p>Les objectifs du stagiaire sont les suivants :</p> <ul style="list-style-type: none"><li>Établir un état de l'art des méthodes développées par les différentes équipes du challenge</li><li>Établir un état de l'art des technologies mêlant réseaux bayésiens et séries temporelles</li><li>Développer et évaluer une méthode à base de réseaux bayésiens sur le challenge de l'ESA</li><li>Une réflexion sur les métriques utilisées</li></ul>

Compétences indispensables	Compétences souhaitées
Deep learning, Python, statistiques	Orbites spatiales, réseaux bayésiens

Observations / Les "+" du stage
<p>Le stagiaire travaillera sur des méthodes avancées d'intelligence artificielle, et intégrera une équipe technique spécialisée en IA couvrant un large périmètre : traitement des langues, image, signal, robotique, cyber, etc.</p>

# ASC2-2024-IA3D-12- Optimisation automatique de lois de pilotage par approche IA

Niveau requis	Durée	Mots-clés
Bac+5	6 mois	automatisme, optimisation, loi de contrôles, apprentissage par renforcement, systèmes complexes

Description du stage
<p>Une des problématiques posées par les systèmes de contrôle haut-niveau est l'interaction et l'interdépendance entre différents algorithmes de contrôle. Par exemple, une amélioration ou un arbitrage effectué sur un contrôleur au niveau d'un drone peut impacter les ordres possibles, on souhaiterait donc pouvoir les optimiser selon des critères plus haut niveau avec un coût de développement minimal, en déléguant un maximum d'étapes d'optimisation.</p> <p>Le sujet de ce stage consiste donc à s'intéresser dans un premier temps à la possibilité d'automatiser l'optimisation de contrôleurs, en utilisant des algorithmes d'optimisation d'hyper-paramètres (processus gaussiens, MCMC...), pour ensuite rendre compte de l'efficacité comparée à d'autres méthodes ou en complément de celles-ci.</p> <p>Les étapes prévisionnelles du stage sont les suivantes :</p> <ul style="list-style-type: none"><li>• Implémentation ou utilisation d'un modèle de simulation de drone (comme le x8, cf. références),</li><li>• Implémentation d'un premier modèle d'autopilotage d'attitude, puis implémentation d'outils et de méthodes pour optimiser le contrôleur,</li><li>• Extension de cette méthode à des contrôleurs plus complexes, afin de faire évoluer/généraliser/raffiner les outils.</li></ul> <p>Enfin, d'autres sujets peuvent être proposés à l'initiative du stagiaire pour explorer de nouvelles méthodes de contrôle et d'optimisation.</p> <p>Étant donné l'étendue des domaines traversés par le stage, il n'est pas demandé au stagiaire d'être instruit sur tous les sujets, mais de manifester une curiosité et une capacité à formaliser les fonctions des différents outils.</p> <p>Enfin, si le stage pourra s'effectuer en Python ou Matlab, une préférence sera manifestée pour des candidats intéressés par l'utilisation de Julia afin de rendre compte de l'environnement technique associé et de ses avantages.</p> <p>Références :</p> <ol style="list-style-type: none"><li>1. K. Gryte, R. Hann, M. Alam, J. Roháč, T. A. Johansen and T. I. Fossen, Aerodynamic modeling of the Skywalker X8 Fixed-Wing Unmanned Aerial Vehicle, ICUAS 2018</li><li>2. <a href="https://github.com/eivindeb/pyfly">https://github.com/eivindeb/pyfly</a> (répertoire implémentant le modèle précédent)</li><li>3. <a href="https://stor-i.github.io/GaussianProcesses.jl/latest/Regression/">https://stor-i.github.io/GaussianProcesses.jl/latest/Regression/</a>, sklearn.GaussianProcessRegressor</li><li>4. <a href="https://github.com/utiasDSL/safe-control-gym">https://github.com/utiasDSL/safe-control-gym</a></li><li>5. <a href="https://members.loria.fr/jbmouret/qd.html">https://members.loria.fr/jbmouret/qd.html</a></li><li>6. Manuel Baltieri, A Bayesian perspective on classical control</li></ol>

Compétences indispensables	Compétences souhaitées
Automatisme, Optimisation, Lois de contrôle	Apprentissage par renforcement, Réseaux de neurones, Recherche opérationnelle

Observations / Les "+" du stage
Le stagiaire partira de travaux préliminaires effectués en interne et commencera dans un cadre défini au début de son stage, mais pourra ensuite, par exemple à travers les références indiquées, proposer ses idées et les mettre en œuvre.



