

## Conception d'une méthodologie d'annotation de données du Naval de Défense

<b>Type de poste :</b>	Thèse de doctorat en sciences des données
<b>Mots clés :</b>	Aide à l'étiquetage des données, processus de décision, apprentissage et transférabilité de la représentation d'expertise, science des données, développement logiciel, naval de défense, facteurs organisationnels et humains.
<b>Etablissements :</b>	IMT Atlantique (Plouzané), Ecole Navale (Lanvéoc), Naval Group (Toulon),
<b>Laboratoires d'accueil :</b>	Lab-Sticc CNRS UMR 6285, EA IRENav
<b>Durée du contrat :</b>	CDD 3 ans – Contrat CIFRE
<b>Etat du poste :</b>	Vacant, ouvert à partir 1 <sup>er</sup> octobre 2024
<b>Spécificité du poste :</b>	Les travaux de ce poste sont susceptibles d'être conditionnés par l'obtention d'une habilitation de la défense nationale

### Contexte du projet de recherche :

Dans le domaine du Naval de Défense, l'annotation des données est une tâche fastidieuse et complexe, impactant toute la chaîne de post-traitement des données : dans le cadre de conception d'algorithmes d'aide à la décision à base d'IA, toute erreur d'étiquetage de l'opérateur humain entraînera des erreurs de classification et dégradera les performances de ces algorithmes d'aide à la décision.

Dans le cadre de la Chaire industrielle NAI ADE<sup>1</sup> (NAlisation de l'Intelligence Artificielle pour l'aide à la DEcision), l'objectif est donc d'augmenter la confiance du marin en des technologies de pointe pouvant faire appel à de l'IA, en optimisant le processus d'annotation, afin que le modèle d'IA résultant de l'apprentissage des annotations du marin soit le reflet de ses connaissances.

En s'appuyant sur les techniques issues des Facteurs Organisationnels et Humains (FOH), de la prise d'expertise et de la science des données, comment optimiser le processus d'annotation des données du Naval de Défense, et ce, en maximisant la qualité et la fiabilité d'annotation tout en minimisant l'effort humain ?

L'objectif de la thèse est d'apporter des réponses à la problématique de l'**optimisation du processus d'annotation des données** du Naval de Défense avec des solutions issues de la science des données.

A ce titre, la personne recrutée pourra utiliser les résultats issus des Facteurs Organisationnels et Humains et outils logiciels obtenus au titre d'études réalisées dans le cadre de la Chaire NAI ADE, et aura l'opportunité d'expérimenter avec les utilisateurs finaux le bienfondé de ses réalisations.

Le/la doctorant(e) portera un intérêt particulier à la complémentarité des recommandations Facteurs Humains avec des solutions issues de la science des données à déterminer pour répondre de façon innovante aux besoins des opérateurs relevés en amont.

### Etat de l'art :

Le sujet inclut notamment les points suivants :

- Réaliser l'analyse de l'activité des opérateurs humains en lien avec l'étiquetage de la donnée [1, 2] et comprendre les profils d'opérateurs hétérogènes étudiés, en interaction avec les spécialistes FOH de la Chaire NAI ADE.
- Identifier les conditions et moyens impactant la qualité d'étiquetage de la donnée [3, 4].

<sup>1</sup> <https://www.imt-atlantique.fr/fr/recherche-innovation/collaborer/chaire/chaire-naiade>

- Etudier l'étiquetage de données par apprentissage semi-supervisé et supervisé [5, 6].
- Etudier l'aide à l'étiquetage à partir d'un historique de données étiquetées [7, 8].
- Caractériser les incertitudes des intervalles de prédiction de l'appartenance d'une donnée à une classe spécifique [9, 10].
- Proposer des pistes d'amélioration des intervalles de certitude de prédiction [11, 12].

L'originalité de ce travail réside dans l'étude de cas d'usage réels où une activité critique d'annotation est réalisée par un opérateur expert humain [13, 14].

### **Question de recherche :**

Dans le cadre des travaux de recherche effectués auparavant, les solutions proposées ne permettent toujours pas d'étiqueter la donnée en fonction de profils d'opérateurs hétérogènes. A l'égard de cette limitation considérable, la question de recherche s'intéresse à la problématique suivante : **quelles méthodes d'aide à la décision pour l'étiquetage de la donnée basées sur l'intelligence artificielle (IA) permettent d'aider à l'annotation à partir d'un historique de données étiquetées par des experts dont le profil est hétérogène, conformément à des conditions de certitude ?**

### **Organisation des travaux de recherche :**

L'objet des travaux sera l'analyse de l'activité des opérateurs humains avec les objectifs suivants :

- Identification des conditions et moyens de la qualité de l'étiquetage et de son amélioration par la prise en compte de l'activité de l'opérateur lors de cette phase (connaissances mises en jeu, critères pris en compte) et par la prise d'expertise croisée de plusieurs opérateurs experts. Approche hybride quantitative/qualitative venant des sciences humaines et sociales.
- Aide à l'annotation à partir d'un historique de données étiquetées : production d'intervalles de confiance de prédiction. Apprentissage automatique.
- Pour les situations où l'intervalle de confiance de prédiction est faible, le second volet des recherches consisterait à identifier les conditions et les moyens d'améliorer le mécanisme de prise de décision en y transférant des éléments de la prise d'expertise des opérateurs.

### **Étape 1 - Étude du processus de décision d'étiquetage**

L'activité des opérateurs étant par nature dépendante du contexte, il s'agira dans un premier temps de concevoir, expérimenter et documenter des recommandations issues des FOH et de la science des données propres à chaque cas d'usage.

Tâche 1 : Etat de l'art sur les solutions d'aide à la décision pour l'étiquetage de la donnée à base d'IA proposées dans la littérature, compréhension de la démarche cognitive et de l'expertise nécessaire pour étiqueter la donnée en vue de caractériser le processus de décision.

Tâche 2 : Analyse des approches et des métriques d'évaluation de la qualité des données à étiqueter et de la performance de l'étiquetage.

Tâche 3 : Identification et validation des éléments et paramètres clés qui facilitent l'étiquetage de données selon le profil de l'opérateur en relation avec l'objectif de l'étiquetage. Proposition de différents modèles d'aide à la décision à base d'IA pour l'étiquetage de la donnée et évaluation de leur performance.

Tâche 4 : Définition des scénarii d'usage, de collecte des données, d'expérimentation et d'analyse des comportements des experts. Test des modules développés dans la Tâche 3 et démonstration des résultats.

Tâche 5 : Publication des résultats dans une conférence et une revue scientifique.

## ***Étape 2 - Étude de la transférabilité du processus de décision***

Application des résultats de l'Étape 1 sur un cas d'usage dans le contexte du naval militaire et un autre cas d'usage selon la disponibilité des données et des experts, afin d'identifier les aspects communs du processus de décision et leur transférabilité d'un cas vers un autre.

Tâche 1 : Analyse de l'intégration de l'expertise croisée de plusieurs opérateurs experts pour l'étiquetage de la donnée.

Tâche 2 : Etat de l'art sur les solutions d'aide à la décision pour l'étiquetage de la donnée à base d'IA utilisant des approches de transférabilité d'expertise. Analyse de l'adéquation des métriques d'évaluation de la qualité des données et de la performance de l'étiquetage, dans le cadre de la transférabilité d'expertise.

Tâche 3 : Définition du processus de décision basé sur la transférabilité d'expertise. Proposition de différents modèles d'aide à la décision à base d'IA pour l'étiquetage de la donnée fondés sur la transférabilité d'expertise.

Tâche 4 : Définition des scénarii d'usage, de collecte des données et d'expérimentation. Test des modèles proposés dans la Tâche 3 et démonstration des résultats.

Tâche 5 : Publication des résultats dans une conférence et une revue scientifique.

## ***Étape 3 - Finalisation***

Tâche 1 : Rédaction du manuscrit de thèse.

Tâche 2 : Préparation de la soutenance.

Tâche 3 : Valorisation et transfert des résultats au sein de l'entreprise.

## ***Encadrement :***

### **John PUENTES**

Professeur  
Département Science des  
Données  
IMT Atlantique

### **Cyril RAY**

Maître de Conférences, HDR  
Institut de Recherche de  
l'Ecole Navale  
Ecole Navale

### **Fabien CHAILLAN**

Data Scientist  
Service Science des données  
Navales  
Naval Group RESEARCH

### **Références:**

- [1] Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., He, L. A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, 2022, 135, pp. 364-381.
- [2] Momber, A. W., Langenkämper, D., Möller, T., Nattkemper, T. W. The exploration and annotation of large amounts of visual inspection data for protective coating systems on stationary marine steel structures. *Ocean Engineering*, 2023, 278, 114337.
- [3] Gomes-Pereira, J. N., Auger, V., Beisiegel, K., Benjamin, R., Bergmann, M., Bowden, D., et al. Current and future trends in marine image annotation software. *Progress in Oceanography*, 2016, 149, pp. 106-120.
- [4] Sexton, T., Brundage, M. P., Hoffman, M., Morris, K. C. Hybrid datafication of maintenance logs from AI-assisted human tags. *Proceedings IEEE International Conference on Big Data*, 2017, pp. 1769-1777.
- [5] Le, H. T., Phung, S. L., Chapple, P. B., Bouzerdoum, A., Ritz, C. H., Chung, L. T. Deep Gabor neural network for automatic detection of mine-like objects in sonar imagery. *IEEE Access*, 2020, 8, 94126-94139. DOI: 10.1109/ACCESS.2020.2995390.
- [6] Pedro, T., Cintra, J., Garcia, N. and De Sá, A.O. Machine learning for ship detection with radar. *IEEE International Workshop on Technologies for Defense and Security*, Rome, Italy, November 2023, pp. 12-17.
- [7] Zhang, W., Yang, D., Wang, H. Data-driven methods for predictive maintenance of industrial equipment: A survey. *IEEE Systems Journal*, 2019, 13(3), pp. 2213-2227.
- [8] Ding, P., Jia, M. Mechatronics equipment performance degradation assessment using limited and unlabeled data, *IEEE Transactions on Industrial Informatics*, 2022, vol. 18, no. 4, pp. 2374-2385.
- [9] Wu, M., Brandhorst, H., Marinescu, M. C., Lopez, J. M., Hlava, M., Busch, J. Automated metadata annotation: What is and is not possible with machine learning. *Data Intelligence*, 2023, 5(1), pp. 122-138.
- [10] Collins, K. M., Barker, M., Espinosa Zarlenga, M., Raman, N., Bhatt, U., Jamnik, M., Sucholutsky, I., Weller, A., Dvijotham, K. Human uncertainty in concept-based AI systems. *Proceedings AAAI/ACM Conference on AI, Ethics, and Society*, 2023, pp. 869-889.
- [11] Zhou, J.T., Fang, M., Zhang, H., Gong, C., Peng, X., Cao, Z. and Goh, R.S.M. Learning with annotation of various degrees. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(9), pp.2794-2804.
- [12] Li, J., Wong, Y., Zhao, Q., Kankanhalli, M.S. Learning to learn from noisy labeled data. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5051-5059.
- [13] Ara, Z., Salemi, H., Hong, S. R., Senarath, Y., Peterson, S., Hughes, A. L., Purohit, H. Closing the knowledge gap in designing data annotation interfaces for AI-powered disaster management analytic systems. *Proceedings 29th International Conference on Intelligent User Interfaces*, 2024, pp. 405-418.
- [14] Watson, E., Viana, T., Zhang, S. Machine Learning Driven Developments in Behavioral Annotation: A Recent Historical Review. *International Journal of Social Robotics*, 2024, pp.1-14.