

Defence-related Research Action - DEFRA

ACRONYME: AHOI

Titre : Adaptive Human Operator Interaction with Autonomous Systems

Durée du projet: 31/12/2023 -01/03/2027

Budget: 1.677.000€

dont contribution IRSD: 1.588.493€

Mots-clés: Explainable AI (XAI), Human-Machine Teaming, Situational Awareness, Decision-Making under uncertainty, Intelligent Automation Systems, Human-centred design.

DESCRIPTION DU PROJET

Dans le domaine des systèmes autonomes, le rôle de la supervision et de l'interaction humaines reste crucial. Les machines, bien qu'habiles à naviguer dans des scénarios incertains, manquent souvent de prendre en compte les considérations éthiques inhérentes aux situations réelles. Cela est particulièrement évident dans les systèmes pilotés par l'IA, car les approches modernes de l'apprentissage automatique ont tendance à être des "boîtes noires", masquant les mécanismes sous-jacents de leurs processus décisionnels. Un risque majeur lié au déploiement généralisé de tels systèmes est que leurs décisions peuvent manquer totalement de transparence, et de ce fait, les humains qui travaillent à leurs côtés, particulièrement dans des scénarios hautement incertains et/ou dangereux, peuvent avoir du mal à surveiller leur performance, à comprendre leurs processus et à déterminer s'ils remplissent leur objectif prévu et fonctionnent dans les limites du système de valeurs socio-éthiques. Tous ces problèmes peuvent être exacerbés par les biais et les heuristiques humains, qui peuvent amener un navigateur à (1) avoir trop confiance en ses compétences et connaissances de navigation et ignorer ou négliger les signaux d'avertissement d'un système autonome (2) rejeter les recommandations du système autonome qui contredisent leurs propres croyances ou hypothèses (par exemple, le biais de confirmation). Le résultat de ce processus est une perte de confiance dans les recommandations du système.

Il y a donc un besoin fort d'investiguer comment l'explicabilité de tels algorithmes et la confiance humaine dans les systèmes se façonnent mutuellement. Pour atteindre cet objectif, le consortium Adaptive Human Operator Interaction with Autonomous Systems (AHOI) réunit une équipe de chercheurs de divers domaines de recherche (Intelligence Artificielle, IA Explicable, science comportementale, formation du personnel maritime et interaction homme-machine) pour enquêter sur les facteurs qui facilitent ou entravent le transfert de responsabilités (points de passation) entre

les systèmes humains et autonomes. Spécifiquement, nous nous concentrerons sur le rôle de la confiance, et comment elle est affectée par le partage d'informations et la transparence, en suivant dynamiquement les indicateurs clés de performance du système et de son opérateur humain dans un contexte maritime opérationnel. Le domaine maritime présente des défis uniques pour les systèmes de navigation autonomes, nécessitant une résilience face à une gamme de contraintes internes et externes. Notre recherche vise à relever ces défis en :

1. Système de Navigation Robuste

Développer un système de navigation autonome résilient qui fonctionne de manière transparente dans des environnements maritimes dynamiques, incluant des défaillances de capteurs, des dysfonctionnements moteur, un trafic dense et des conditions météorologiques défavorables. Ce système utilisera des techniques avancées d'apprentissage automatique pour se généraliser à des paramètres inconnus et non structurés avec une reconfiguration minimale.

2. IA Explicable pour des Décisions Fiables

Étudier l'impact des méthodes d'explication de modèle (XAI) sur la confiance et les processus de prise de décision des opérateurs humains. Nous examinerons comment le XAI peut améliorer la compréhension humaine des décisions générées par l'IA, en particulier pour les opérateurs ayant des niveaux d'expérience et d'expertise variés.

3. Collaboration Homme-Machine dans la Prise de Décision

Explorer la relation entre les biais humains et la perception de la transparence dans l'interaction homme-machine (HMI). Nous mènerons des expériences pour déterminer le point de contact optimal entre les humains et les machines, leur permettant de collaborer efficacement dans des situations à haute incertitude ou biais.

4. HMI Avancé pour une Transparence Accrue

Concevoir un outil HMI avancé qui présente non seulement les décisions générées par l'IA, mais fournit également des aperçus de leur raisonnement sous-jacent. Ce HMI ajustera dynamiquement sa sortie en fonction des profils d'utilisateur, en adaptant les explications aux niveaux d'expérience individuels et en favorisant la confiance dans la prise de décision de l'IA.

5. XAI Adaptatif et Optimisation Basée sur le Feedback

Tirer parti des nouvelles méthodologies en XAI et des logiciels de visualisation pour créer un HMI dynamique et interactif. Cet HMI recueillera les retours des utilisateurs et adaptera les explications de XAI en conséquence, améliorant la performance de l'IA et favorisant un cycle d'apprentissage continu.

Notre approche holistique de la navigation autonome vise à créer un système à la fois robuste et transparent, permettant une collaboration homme-machine sans heurts dans le domaine maritime complexe. En s'attaquant aux défis de la fiabilité et de l'explicabilité, nous pouvons ouvrir la voie à des opérations autonomes sûres et efficaces à l'avenir.

Ce travail est pertinent pour diverses applications liées à la défense, où les systèmes peuvent fonctionner indépendamment, mais nécessitent toujours l'implication humaine et la bonne chaîne de commandement. Bien que centrée sur le scénario de navigation de vaisseaux autonomes, nos résultats de recherche peuvent être généralisés à des applications telles que la chasse aux mines, les missions d'Intelligence, Surveillance et Reconnaissance (ISR), les opérations UAV et UGV, etc. Dans AHOI, iMec mènera des recherches sur l'IA et le XAI, tandis que l'UA et l'AMA réaliseront une étude approfondie des biais et du profilage des opérateurs humains. MAHI mènera des recherches sur la conscience situationnelle des navires autonomes et concevra le HMI final.

COORDONNÉES

Coordinateur

Dirk Van Rooy
University of Antwerp, Product development.
e-mail: Dirk.VanRooy@uantwerpen.be

Partenaires

Ali Anwar, José Oramas
IMEC - Internet Data Lab (IDLab) Antwerp
e-mail: Ali.Anwar@imec.be , Jose.Oramas@imec.be

Pieter-Jan Note
MAHI
e-mail: pieterjan.note@mahi.be

Pieter Maes
Antwerp Maritime Academy
e-mail: pieter.maes@hzs.be

LIEN(S) DU PROJET

/