Defence-related Research Action - DEFRA

ACRONIEM: AHOI

Titel: Adaptive Human Operator Interaction with Autonomous Systems

Duur van het project: 31/12/2023- 01/03/2027

Totaal budget: 1.667.000€

waarvan bijdrage KHID: 1.588.493€

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

BESCHRIJVING VAN HET PROJECT

In het domein van autonome systemen blijft de rol van menselijk toezicht en interactie cruciaal. Machines zijn weliswaar bedreven in het navigeren door onzekere scenario's, maar schieten vaak tekort in het meenemen van ethische overwegingen die in echte situaties ingebed zijn. Dit is met name duidelijk bij AI-gestuurde systemen, aangezien moderne machine learning benaderingen vaak "black boxes" zijn, die de onderliggende mechanismen van hun besluitvormingsprocessen verduisteren. Een primair risico bij de toepassing van dergelijke systemen is dat hun beslissingen volledig aan transparantie kunnen ontbreken, waardoor menselijke gebruikers, zeker in onzekere en/of onveilige scenario's, moeite hebben met het monitoren van hun prestaties, het begrijpen van hun processen en het bepalen of ze hun beoogde doel vervullen en opereren binnen de grenzen van het socio-ethische waardesysteem. Al deze kwesties kunnen verergerd worden door menselijke vooroordelen en heuristieken, die een navigator kunnen doen (1) overmoedig zijn in hun navigatievaardigheden en -kennis, en waarschuwingssignalen van een autonoom systeem negeren of over het hoofd zien (2) de aanbevelingen van het autonome systeem afwijzen die in strijd zijn met hun eigen overtuigingen of aannames (bv bevestigingsbias). Het resultaat van dit proces is een afbraak van vertrouwen in de aanbevelingen van het systeem.

Er is dus een sterke behoefte om te onderzoeken hoe de uitlegbaarheid van dergelijke algoritmen en het menselijk vertrouwen in systemen elkaar vormgeven. Om dit te bereiken brengt het consortium Adaptive Human Operator Interaction with Autonomous Systems (AHOI) een team van onderzoekers samen uit verschillende onderzoeksgebieden (Kunstmatige Intelligentie, Uitlegbare AI, gedragswetenschap, maritieme personeelstraining en mens-machine interactie) om de factoren te onderzoeken die de overdracht van verantwoordelijkheden (overdrachtspunten) tussen mens en autonome systemen vergemakkelijken of belemmeren. Specifiek zullen we ons richten op de rol van

vertrouwen en hoe dit wordt beïnvloed door informatie-uitwisseling en transparantie, door dynamisch de belangrijkste prestatie-indicatoren van het systeem en zijn menselijke operator te volgen binnen een operationele, maritieme context. De maritieme sector presenteert unieke uitdagingen voor autonome navigatiesystemen, die veerkracht vereisen tegen een reeks interne en externe beperkingen. Ons onderzoek is gericht op het aanpakken van deze uitdagingen door:

1. Robuust Navigatiesysteem

Het ontwikkelen van een veerkrachtig autonoom navigatiesysteem dat naadloos functioneert in dynamische maritieme omgevingen, inclusief sensorstoringen, motorstoringen, druk verkeer en ongunstige weersomstandigheden. Dit systeem zal geavanceerde machine learning technieken gebruiken om zich te generaliseren naar onbekende, ongestructureerde omgevingen met minimale herconfiguratie.

2. Uitlegbare AI voor Betrouwbare Beslissingen

Het onderzoeken van de impact van modelverklaringsmethoden (XAI) op het vertrouwen en de besluitvormingsprocessen van menselijke operators. We zullen onderzoeken hoe XAI het menselijk begrip van door AI gegenereerde beslissingen kan verbeteren, met name voor operators met verschillende ervarings- en deskundigheidsniveaus.

3. Mens-Machine Samenwerking in Besluitvorming

Het verkennen van de relatie tussen menselijke vooroordelen en de perceptie van transparantie in de interactie tussen mens en machine (HMI). We zullen experimenten uitvoeren om het optimale contactpunt tussen mensen en machines te bepalen, waardoor ze effectief kunnen samenwerken in situaties met hoge onzekerheid.

4. Geavanceerde HMI voor Verhoogde Transparantie

Het ontwerpen van een geavanceerd HMI dat niet alleen door AI gegenereerde beslissingen presenteert, maar ook inzicht geeft in de onderliggende redenering. Deze HMI zal dynamisch zijn uitvoer aanpassen op basis van gebruikersprofielen, verklaringen op maat bieden voor individuele ervaringsniveaus en vertrouwen in AI-besluitvorming bevorderen.

5. Adaptieve XAI en Feedback-Gedreven Optimalisatie

Het benutten van nieuwe methodologieën in XAI en visualisatiesoftware om een dynamische en interactieve HMI te creëren. Deze HMI zal gebruikersfeedback verzamelen en XAI-verklaringen dienovereenkomstig aanpassen, de AI-prestaties verbeteren en een continue leerproces bevorderen.

Onze holistische benadering van autonome navigatie streeft ernaar een systeem te creëren dat zowel robuust als transparant is, waardoor naadloze mens-machine samenwerking in het complexe maritieme domein mogelijk is. Door de uitdagingen van betrouwbaarheid en uitlegbaarheid aan te pakken, kunnen we de weg vrijmaken voor veilige en efficiënte autonome operaties in de toekomst.

Dit werk is relevant voor verschillende defensiegerelateerde toepassingen, waar systemen mogelijk onafhankelijk opereren, maar nog steeds menselijke betrokkenheid en de juiste bevelsstructuur vereisen. Hoewel ons onderzoek zich toespitst op het scenario van autonome scheepsnavigatie, kunnen onze onderzoeksresultaten worden gegeneraliseerd naar toepassingen zoals mijnenjacht, Intelligence, Surveillance and Reconnaissance (ISR)-missies, UAV-operaties en UGV-operaties enz. In AHOI zal iMec onderzoek doen naar AI en XAI, terwijl UA en AMA een diepgaande studie zullen uitvoeren naar de vooroordelen en profilering van menselijke operators. MAHI zal onderzoek doen naar het situationeel bewustzijn van de autonome schepen en de uiteindelijke HMI ontwerpen.

CONTACTINFORMATIE

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

Partners
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

LINK(S) NAAR PROJECT

/