Concevoir des explications d'IA appropriables par les opérateurs : une démarche prospective appliquée à la patrouille maritime

MOURGAUD Luca, BOBILLIER-CHAUMON Marc-Eric, GRISVARD Olivier, LEMAITRE Denis

Identification des cas d'usages

Les progrès des capteurs et des moyens de communication ont conduit, dans l'ensemble des milieux d'engagement (surface, sous-marin, aérien), à une densification continue des pistes traitées, c'est-à-dire des contacts détectés et suivis par les capteurs de la plateforme. À bord d'un avion de patrouille maritime (ATL2), l'opérateur radar est chargé de détecter, classifier et identifier ces pistes, puis de transmettre au coordinateur tactique (TACCO) les éléments nécessaires à la conduite de la mission. L'augmentation du nombre de contacts, la diversité des sources d'information (radar, AIS, guerre électronique, imagerie) et la contrainte de temps sur laquelle repose la décision rendent cette activité de plus en plus complexe et exigeante. C'est dans ce contexte que se déploient des systèmes d'aide à la décision fondés sur l'Intelligence Artificielle (IA) : des systèmes qui, à partir des données produites par les capteurs, formulent des recommandations de classification et signalent des situations anormales (par exemple des incohérences entre l'AIS déclaré et ce que détecte le radar). L'opérateur conserve cependant la charge d'évaluer la recommandation, de la discuter au sein de l'équipage et, in fine, de la valider ou de l'écarter. Le cas d'usage étudié ici est celui d'un système d'IA d'aide à la décision, actuellement en phase de conception, destiné aux opérateurs radar de l'ATL2.

Le problème opérationnel se situe dans les conditions de mobilisation de ces recommandations. Les environnements militaires se caractérisent par une forte pression temporelle, une organisation collective hiérarchisée et des exigences élevées de traçabilité des décisions (Schraagen, 2023) : les informations produites par le système ne sont pas seulement destinées à la compréhension individuelle d'un opérateur, elles doivent circuler dans les échanges qui structurent la conduite de la mission et permettre de rendre compte des décisions auprès de la chaîne de commandement. Or les travaux en Explainable Artificial Intelligence (XAI) demeurent, dans leur ensemble, technocentrés. L'explication y est principalement construite à partir du fonctionnement interne des modèles. Bennani, Fréjus et Bobillier-Chaumon (2025) défendent, dans le cadre de ces limites, une perspective située et incarnée par et pour les métiers. Nous inscrivons notre démarche dans cette approche, en proposant que les caractéristiques mêmes de l'explication (contenu, format, destinataires) soient dérivées de l'analyse de l'activité de travail. Nous qualifions cette approche d'explicabilité située.

L'objectif, *in fine*, est une acceptation et une appropriation effective de ces systèmes par les équipages : des recommandations dont les justifications soient exploitables dans les différents tempos de mission, transmissibles au sein de l'équipage et interrogeables par la hiérarchie. La démarche est transposable à d'autres contextes de décision coopérative sous contrainte temporelle, militaires ou civils (contrôle aérien, supervision de systèmes critiques). Un protocole expérimental permettant la comparaison de plusieurs modalités d'explication lors de tâches de classification de contacts maritimes sous différentes contraintes temporelles est en cours de préparation.

Approche technique proposée

L'approche proposée est une méthode de conception d'explications pour systèmes d'IA d'aide à la décision. Elle ne se substitue pas aux techniques développées en XAI, qui répondent à la

question de la fidélité de l'explication au modèle, mais les complémente en adressant une autre question : à quelles conditions cette explication devient-elle utilisable dans l'activité réelle des opérateurs ? Le verrou traité est celui de l'utilisabilité opérationnelle des explications produites par un système d'IA en environnement contraint.

La méthode relève de la psychologie du travail et de l'analyse de l'activité. Elle repose sur deux propositions théoriques : un artefact ne devient un instrument utile qu'à travers son appropriation dans l'activité (Rabardel, 1995), et l'acceptation d'une technologie se construit dans l'usage effectif plutôt qu'en amont de celui-ci (Bobillier-Chaumon, 2016). Le système étudié n'étant pas encore déployé, ces propositions sont mobilisées de manière prospective : les dimensions de l'acceptation située servent de grille de lecture pour analyser l'activité actuelle des utilisateurs du futur système, et pour en dériver les conditions d'une intégration acceptable et appropriable du futur système (Cippelletti, Adelé, Dionisio & Bobillier-Chaumon, 2023). Il en découle qu'une explication ne peut être conçue indépendamment de l'activité dans laquelle elle sera mobilisée, et que sa conception passe par quatre étapes :

(1) analyser l'activité réelle des utilisateurs du système à concevoir ;

(2) identifier les dimensions de cette activité qui structurent l'usage possible des explications ;

(3) dériver de ces dimensions des principes de conception de l'explication ;

(4) confronter ces principes à des données comportementales sur simulateur.

Les trois premières étapes ont été mises en œuvre sur le cas d'usage. Quatorze entretiens semi-directifs conduits auprès d'opérateurs en activité, combinant instruction au sosie revisitée (Clot, 2008) et technique des incidents critiques (Flanagan, 1954), ont été analysés par analyse thématique réflexive (Braun & Clarke, 2021) selon une démarche abductive. Trois dimensions de l'activité structurent les attentes exprimées vis-à-vis d'une future explication : sa temporalité (une même explication ne peut être exploitée de la même manière en phase de tempo calme et en phase de tempo élevé), son organisation collective (les informations circulent horizontalement entre opérateurs spécialisés et verticalement vers le coordinateur tactique, avec des attentes différentes selon la position), et le développement de l'expertise (à poste équivalent, les opérateurs expérimentés et moins expérimentés attribuent des fonctions différentes à une même explication).

Ces trois dimensions se traduisent en trois principes de conception directement mobilisables pour le développement du futur système : un niveau de détail ajusté au tempo de la mission, un contenu différencié selon la position dans la division du travail, un format permettant la reformulation et la transmission au sein de l'équipage plutôt que la seule consultation individuelle.

Le niveau de maturité est celui d'un concept étayé par une analyse de terrain approfondie auprès des utilisateurs du futur système. La quatrième étape est en cours de préparation : un protocole expérimental comparera plusieurs modalités d'explication, dont une explication de type XAI standard et une explication conçue selon les principes issus de l'analyse présentée ici, lors de tâches de classification de contacts maritimes sous différentes contraintes temporelles.

Bibliographie

Bennani, R., Fréjus, M., & Bobillier Chaumon, M.-E. (2025). *De l'explicabilité à l'explication des IA : perspective située incarnée par et pour les métiers* [Communication]. Conférence EGC, Atelier EXPLAIN'AI 2025, Strasbourg, France. https://hal.science/hal-04920628

Bobillier Chaumon, M.-E. (2016). L'acceptation située des technologies dans et par l'activité : premiers étayages pour une clinique de l'usage. *Psychologie du Travail et des Organisations, 22*(1), 4–21. https://doi.org/10.1016/j.pto.2016.01.001

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Cippelletti, E., Adelé, S., Dionisio, C., & Bobillier Chaumon, M.-E. (2023). Prospecter l'acceptation située d'une technologie en contexte professionnel, une étude qualitative exploratoire sur l'exploitation de la route. *Psychologie du Travail et des Organisations, 29*(4), 249–271. https://doi.org/10.1016/j.pto.2023.09.001

Clot, Y. (2008). *Travail et pouvoir d'agir*. Presses Universitaires de France.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327–358. https://doi.org/10.1037/h0061470

Rabardel, P. (1995). *Les hommes et les technologies : approche cognitive des instruments contemporains*. Armand Colin.

Schraagen, J. M. (2023). Responsible use of AI in military systems: Prospects and challenges. *Ergonomics, 66*(11), 1719–1729. https://doi.org/10.1080/00140139.2023.2278394