



HAL
open science

Human-Autonomy Teaming: mieux prendre en compte l'état de l'opérateur humain, pour mieux adapter la coopération humain-machine

Philippe Rauffet

► **To cite this version:**

Philippe Rauffet. Human-Autonomy Teaming : mieux prendre en compte l'état de l'opérateur humain, pour mieux adapter la coopération humain-machine. Conférence EPIQUE, Symposium "Avenir et défis de l'Human-Autonomy Teaming dans le domaine de l'aéronautique", Jul 2023, Paris, France. hal-04154628

HAL Id: hal-04154628

<https://hal.science/hal-04154628v1>

Submitted on 27 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Human-Autonomy Teaming : mieux prendre en compte l'état de l'opérateur humain, pour mieux adapter la coopération humain-machine

Philippe Rauffet

Equipe FHOOX, Lab-STICC UMR CNRS 6285, Université Bretagne Sud, philippe.rauffet@univ-ubs.fr

Catégorie de soumission : communication de symposium

RÉSUMÉ

Le paradigme Human-Autonomy Teaming (HAT) a récemment émergé pour concevoir des équipes hybrides, où coopèrent opérateurs humains et agents artificiels autonomes. Un défi majeur du HAT est de transformer ces agents autonomes en de meilleurs coéquipiers, plus interdépendants avec leurs partenaires humains. Afin de répondre à ces nouveaux enjeux, deux axes de recherche sont investigués, notamment au travers d'exemples du domaine aérien. Le premier axe porte sur la classification de la charge mentale et des modes de contrôle de l'opérateur, afin de rendre la machine plus sensible aux difficultés rencontrées par l'humain. Le deuxième axe traite des mécanismes d'adaptation de ces équipes, pour rendre la machine plus "compatible" avec l'homme. Nous nous intéressons notamment à l'adaptation dynamique de la coopération humain-machine, où le système autonome peut être un coéquipier opérationnel, ou un « coach » supervisant l'équipe. Cet article dresse enfin quelques perspectives autour du dialogue humain-machine et de la confiance dans l'autonomie.

MOTS-CLÉS

Human-Autonomy Teaming, Charge mentale, Contrôle cognitif, Automatisation et Interfaces Adaptatives

1 INTRODUCTION : LE PARADIGME HUMAN-AUTONOMY TEAMING (HAT)

Les récentes avancées sur l'intelligence artificielle, les systèmes autonomes et la modélisation cognitive interrogent sous un jour nouveau les processus de prise de décision et la répartition du travail entre les humains et les machines. De nombreuses questions doivent être traitées pour gérer l'activité conjointe et les nouvelles interférences qui apparaissent entre opérateurs humains et systèmes cyber-physiques dotés de capacités accrues. Le paradigme Human-Autonomy Teaming (HAT) a récemment émergé pour caractériser l'évolution de ces équipes hybrides et développer de nouvelles façons de penser les mécanismes de coopération au sein de ces entités (O'Neill et al., 2020). Dans ce paradigme, le système intelligent doit être perçu et accepté par les opérateurs humains comme un membre à part entière de l'équipe, et il doit avoir un rôle et une contribution clairement identifiables, comme cela pourrait être le cas avec un coéquipier humain. L'agent autonome doit donc être considéré comme un membre légitime de l'équipe et non plus comme un simple outil de soutien (Lyons et al., 2018). Cette distinction entre "coéquipier" et "outil" est un élément discriminant, isolant la configuration HAT des configurations précédemment explorées autour des concepts de HAI (Human-Automation Interaction) et de LOA (Levels of Automation). Par conséquent, dans un HAT, le système autonome doit également être en mesure de participer à des activités correspondant au "travail d'équipe", telles que la coordination ou la répartition de l'activité. Comme l'ont souligné Johnson et al. (2011), "l'amélioration de l'efficacité des équipes HAT ne repose pas uniquement sur les travaux visant à rendre les machines encore plus autonomes et plus indépendantes. Un défi important consiste également à trouver des solutions pour transformer les systèmes autonomes en meilleurs coéquipiers, capables de rejoindre les opérateurs humains dans des activités de coopération très interdépendantes et sophistiquées". Abbass et al. (2018) soulignent à nouveau cette perspective en expliquant que les systèmes autonomes doivent mûrir et apprendre à devenir des coéquipiers fiables. Faisant une analogie avec les modèles de développement et

d'émancipation des enfants qui deviennent adultes, ces auteurs expliquent que les machines sont déjà passés de la phase de dépendance (où les humains sont toujours présents pour contrôler leur comportement, pour les passer d'un état ou d'un mode à un autre) à la phase d'indépendance (où les machines, conscientes de l'environnement grâce à leurs capteurs et leurs capacités d'analyse de données, peuvent effectuer de nombreuses tâches et s'adapter à différentes situations sans intervention humaine). Cependant, cette phase d'indépendance est limitée et nécessairement transitoire : la machine, comme Icare se libérant de la tutelle de son père Dédale, ne peut pas continuer à progresser si elle ne connaît pas l'interdépendance sociale. En effet, c'est en interagissant avec les autres que la machine pourrait adapter son comportement, mais aussi influencer le comportement des autres et améliorer les performances collectives. Cependant, la voie vers l'interdépendance est tortueuse : il faut éviter le piège de la codépendance, où les machines et les humains ne peuvent pas se passer l'un de l'autre, et où ils risquent tous les deux de perdre leur autonomie. Comme prévient Koenig (2019), les systèmes intelligents, en assistant trop les humains, pourraient mettre fin à l'idée d'un humain autonome et responsable (avec une perte de compétences et des difficultés à rester dans la boucle de contrôle), qui ne pourrait plus à son tour défier et entraîner la machine à résoudre de nouveaux problèmes.

L'interdépendance des agents autonomes est censée augmenter l'efficacité des équipes homme-autonomie, notamment sur 4 critères (cf. Figure 1, dans les 4 rectangles) : sur la partie « *Team* » en bas à gauche de la figure, il s'agit de jouer sur la composition et l'organisation de l'équipe, avec une amélioration des compétences individuelles permettant une meilleure *qualité dans la prise de décisions* (en permettant aux humains de mieux détecter les faux positifs ou les omissions du diagnostic de l'agent autonome), et une diminution de la *charge mentale* l'opérateur humain (en évitant de lui donnant trop d'informations ou trop de tâche à traiter pour prendre une décision ou interagir dans un temps imparti) ; sur la partie « *Work* » en haut à droite de la figure, en fournissant aux humains une meilleure *conscience de la situation* et une *meilleure calibration de la confiance dans le système autonome* (sur la rationalité ou la fiabilité du comportement de l'agent) pour réaliser les tâches. Pour garantir une coopération humain-machine efficace (cf. Figure 1), il faut équilibrer ces 4 critères, afin d'éviter que l'équipe hybride ne devienne une *équipe aveugle* (avec une conscience de la situation limitée), une *"équipe instable"* (généralisant trop de charge mentale), une *"équipe incompétente"* (où la fiabilité de la machine et/ou la qualité des décisions humaines sont faibles), ou encore une *"pseudo-équipe"* (avec une confiance mal calibrée).

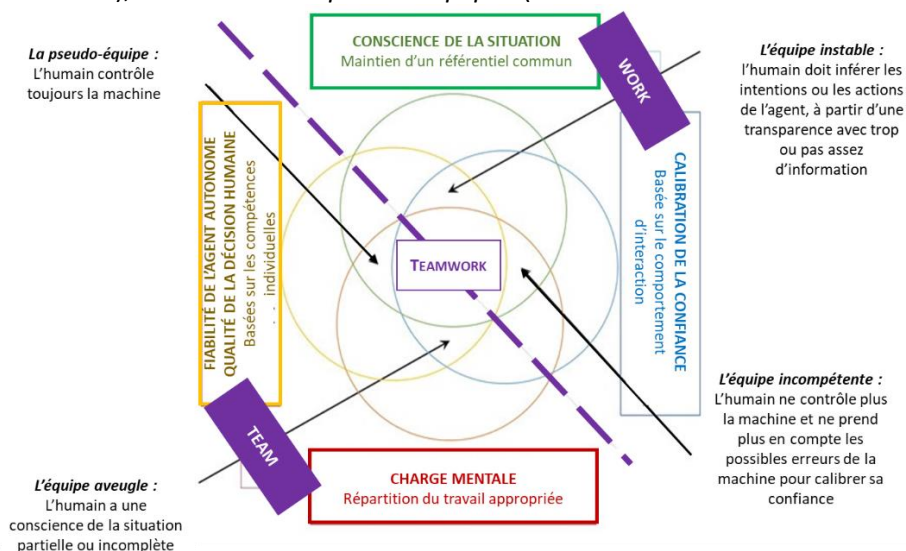


Figure 1 : Dimensions pour optimiser la HAT (Rauffet, 2021, adaptée de d'Angelo et al., 2019)

2 AXES DE RECHERCHE POUR LE HAT ET EXEMPLES APPLICATIFS DANS LE DOMAINE AERIEN

Afin d'améliorer l'interdépendance humain-système, et transformer les machines en meilleurs coéquipiers, deux axes de recherche peuvent particulièrement être investigués (Rauffet et al., 2021).

Axe 1 : Détection – Rendre l’agent autonome plus sensible à l’état cognitif de l’opérateur. Ce premier axe traite de la mesure, de la classification et de la surveillance de certains états cognitifs liés à la performance humaine, tels que la charge mentale, la fatigue mentale ou le stress. L'idée est de doter la machine d'une capacité de détection et de prédiction lorsque l'opérateur humain rencontre des difficultés et n'arrive plus à maîtriser la situation dynamique ou le système autonome. Cette sensibilité à l'état humain pourrait être considérée comme une première étape pour rendre la machine plus interdépendante de l'activité humaine. En effet, la détection d'un état non fonctionnel de l'opérateur pourra conduire à des adaptations du partage du travail entre l'homme et la machine, et cela aidera également les humains à prendre conscience de leurs états actuels et de leurs limites potentielles, grâce à un retour réflexif du système autonome. La littérature est abondante sur la question du suivi des états cognitifs tels que la charge mentale, la fatigue cognitive ou la tunnelisation attentionnelle (Dehais et al., 2017). Dans le cadre de projets en partenariat avec la DGA et Dassault Aviation sur l'activité de supervision d'essais de drones, nous nous sommes intéressés à la création d'un indicateur de charge mentale synthétique qui fusionne des données objectives multidimensionnelles, telles que les exigences dynamiques de la tâche, les signaux neurophysiologiques de l'opérateur, la focalisation attentionnelle, les performances sur la tâche, ainsi qu'un codage objectif des régulations mises en œuvre par l'humain (priorisation de certaines tâches, relâchement de certaines contraintes, etc.). En utilisant la logique floue, des réseaux bayésiens et des techniques de machines learning (SVM et k-NN), nous avons développé des algorithmes permettant une détection multi-classes de 3 à 5 niveaux de charge mentale de l'opérateur (Kostenko et al., 2022). La validité de cette détection a été démontrée en corrélant les valeurs de cet indicateur objectif de charge mentale avec les scores donnés par les participants sur des questionnaires subjectifs (ISA ou SMEQ). Par ailleurs, nous avons exploré plus finement les phénomènes d'engagement et de régulation, lesquels peuvent expliquer certaines variations de charge mentale qui ne sont pas dues aux variations des exigences de la tâche. Au travers de projets avec Thalès et Airbus sur l'activité multitâche des pilotes d'avion, nous avons ainsi opérationnalisé l'évaluation des modes de contrôle cognitif tels que proposés dans le modèle COCOM (Hollnagel, 2003), et identifié un lien significatif entre l'implémentation de certains modes et une activation du cortex préfrontal, mesurée à l'aide d'un capteur fNIRS (Rauffet et al., 2022).

Axe 2 : Adaptation dynamique – Rendre l’agent autonome plus compatible avec l’opérateur. Le deuxième axe se concentre davantage sur les mécanismes d'adaptation de l'équipe humain-système, rendant la machine plus " compatible " avec l'humain. Cet axe peut bien sûr être lié à la première question de recherche puisque l'adaptation peut être conçue, voire déclenchée dynamiquement, grâce à la détection ou à la prédiction d'un état non fonctionnel de l'opérateur. Cette question peut être étudiée dans différentes configurations d'équipes hybrides. D'une part, il existe des organisations où le système autonome est à *l'intérieur de l'équipe* - en tant que coéquipier de l'opérateur humain, il détecte et gère les interférences entre leurs activités respectives. Dans ce cadre, on peut ainsi travailler sur l'amélioration du dialogue humain-machine avec des interfaces de supervision adaptative, ou encore sur l'allocation dynamique des fonctions (Rauffet et al., 2015). Dans nos études sur les activités de supervision de drones (Kostenko et al., 2016), nous avons utilisé notre indicateur synthétique de charge mentale pour déclencher des aides visuelles en cas de forte charge (augmentation de l'interface sur les cibles à surveiller), ou une délégation de certaines tâches en cas de surcharge (automatisation de la localisation précise de la cible). Ces adaptations sur une activité exploratoire ont montré une meilleure efficacité (une meilleure performance, pour un niveau d'effort mental équivalent dans les conditions avec et sans assistance). D'autre part, on peut également étudier les structures où le système autonome est *au-dessus de l'équipe* - comme un coach synthétique supervisant les interférences qui se produisent au sein d'une équipe humaine. C'est par exemple ce qui a été développé dans le cadre du projet ANR ASTRID TAPAS, où une preuve de concept a été proposée pour suivre l'activité de coopération de deux pilotes de chasse de Rafale, avec un suivi des communications et des états physiologiques. Nous avons ainsi implémenté la classification et détection des tâches collectives critiques dans des scénarios de qualification des pilotes, en étudiant la charge mentale générée (mesurée avec différents capteurs, notamment un

EEG, un capteur GSR et oculomètre double) et/ou la dégradation de la communication (en termes d'écart à des standards, d'hésitations, ou de demandes de clarification d'un pilote vers l'autre). Cette détection alimentait ensuite une assistance cognitive, qui pouvait inciter un des pilotes à répéter une information ou à l'enrichir, pour garantir le maintien d'un référentiel commun (Lassalle et al., 2017).

3 PERSPECTIVES ET NOUVEAUX CHAMPS D'INVESTIGATION

Ces deux axes, déjà explorés depuis quelques décennies par la communauté HAI (Human-Automation Interaction), continuent d'être investigués et développés par la communauté HAT, qui doit traiter ces questions en tenant compte de la spécificité des agents autonomes et apprenants. Afin de faire émerger des équipes hybrides effectives et performantes, d'autres pistes attirent également l'attention depuis quelques années. La question de la transparence et de l'explicabilité sont par exemple des éléments de dialogue et d'interactions qu'il faut intégrer (Chen et al., 2017) pour améliorer la confiance entre l'humain et le système. Par ailleurs, comme dans le cadre d'une équipe humaine, la performance s'améliore avec l'expérience et l'apprentissage. Il faut donc réfléchir à des nouvelles modalités d'entraînement de ces équipes hybrides, et ce sur la durée, en travaillant sur l'apprentissage mutuel et l'entraînement adaptatif (Demir et al., 2019).

4 BIBLIOGRAPHIE

- Abbass, H. A., Scholz, J., & Reid, D. J. (2018). *Foundations of trusted autonomy*. Springer Nature.
- Chen, J. Y., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical issues in ergonomics science*, 19(3), 259-282.
- D'Angelo, M., Cervero, R., Durning, S., & Varpio, L. (2019). The teamwork model: Proposing a model for studying interprofessional healthcare teams. *MedEdPublish*, 8.
- Dehais, F., Behrend, J., Peysakhovich, V., Causse, M., & Wickens, C. D. (2017). Pilot flying and pilot monitoring's aircraft state awareness during go-around execution in aviation: A behavioral and eye tracking study. *The International Journal of Aerospace Psychology*, 27(1-2), 15-28.
- Demir, M., McNeese, N. J., Johnson, C., Gorman, J. C., Grimm, D., & Cooke, N. J. (2019, April). Effective team interaction for adaptive training and situation awareness in human-autonomy teaming. In *IEEE conference on cognitive and computational aspects of situation management (CogSIMA)* (pp. 122-126).
- Hollnagel, E. (2003). Context, cognition and control. *Coop Process Manag*, 27-52.
- Johnson, M., Bradshaw, J. M., Feltovich, P., Jonker, C., van Riemsdijk, B., Sierhuis, M. (2011). The fundamental principle of coactive design: Interdependence must shape autonomy. *Coin international workshops, LNAI* (Vol. 6541, pp. 172-191).
- Koenig, G. (2019). *La fin de l'individu. Voyage d'un philosophe au pays de l'intelligence artificielle*. Éditions de l'Observatoire.
- Kostenko, A., Rauffet, P., Moga, S., & Coppin, G. (2019). Operator functional state: measure it with attention intensity and selectivity, explain it with cognitive control. *Third International Symposium, H-WORKLOAD 2019*, November. Rome, Italy
- Kostenko, A., Rauffet, P., & Coppin, G. (2022). Supervised classification of operator functional state based on physiological data: Application to drones swarm piloting. *Frontiers in Psychology*, 12.
- Lassalle, J., Rauffet, P., Leroy, B., Guérin, C., Chauvin, C., Coppin, G., & Saïd, F. (2017). COmmunication and WORKload analyses to study the COLlective WORK of fighter pilots: the COWORK 2 method. *Cognition, Technology & Work*, 19, 477-491.
- Lyons, J. B., Mahoney, S., Wynne, K. T., & Roebke, M. A. (2018). Viewing machines as teammates: A qualitative study. *2018 AAAI Spring Symposium Series*
- O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2020). Human-Autonomy Teaming: A Review and Analysis of the Empirical Literature. *Human Factors*, 0018720820960865.
- Rauffet, P., Chauvin, C., Morel, G., & Berruet, P. (2015). Designing sociotechnical systems: a CWA-based method for dynamic function allocation. In *Proceedings of the European Conference on Cognitive Ergonomics 2015* (pp. 1-8).
- Rauffet, P. (2021). *Tools and methods for Human-Autonomy Teaming: contribution to cognitive state monitoring and system adaptation* (Doctoral dissertation, Université Bretagne Sud).
- Rauffet, P., Saïd, F., Laouar, A., Chauvin, C., & Bressolle, M. C. (2022). Cognitive Control: Transitions in Control Modes Under Different Level of Workload and fNIRS Sensitivity. *CCIS book series, volume 1609*, Springer