Modélisation informatique des erreurs de décisions d'un opérateur humain en aviation ou en médecine

Proposition de stage de fin d'études

Contexte du stage

Ce stage est proposé dans le cadre d'un projet de recherche baptisé **IDEFIX** pour « Intelligence artificielle pour le Désengagement des Erreurs de FIXation », financé par l'Agence Nationale de la Recherche. Ce projet regroupe 6 organismes de recherche nationaux : l'université Paris-Saclay, le CNRS, l'université Lyon-Claude-Bernard, l'université technologique de Compiègne, l'IRBA et l'ONERA. L'objectif de ce projet est d'utiliser l'Intelligence Artificielle et la Réalité Virtuelle & Augmentée pour aider les pilotes d'avion et les médecins à mieux gérer leurs erreurs de fixation, c'est-à-dire le fait d'ignorer des informations qui contredisent leur diagnostic initial. Ces erreurs sont en effet responsables de très nombreux décès chaque année.

Plus d'information sur le projet sur : https://anr-idefix.lisn.upsaclay.fr/

Dans le cadre du projet IDEFIX, nous souhaitons recruter un·e étudiant·e en informatique niveau bac+5 pour son stage de fin d'études, avec l'objectif d'une poursuite en thèse. Le travail de stage porte sur le modèle logique qui servira de base pour la détection des biais cognitifs (dont font partie les erreurs de fixations). Des compétences en **modèles formels** et en **intelligence artificielle fondée sur la logique** sont donc requis.

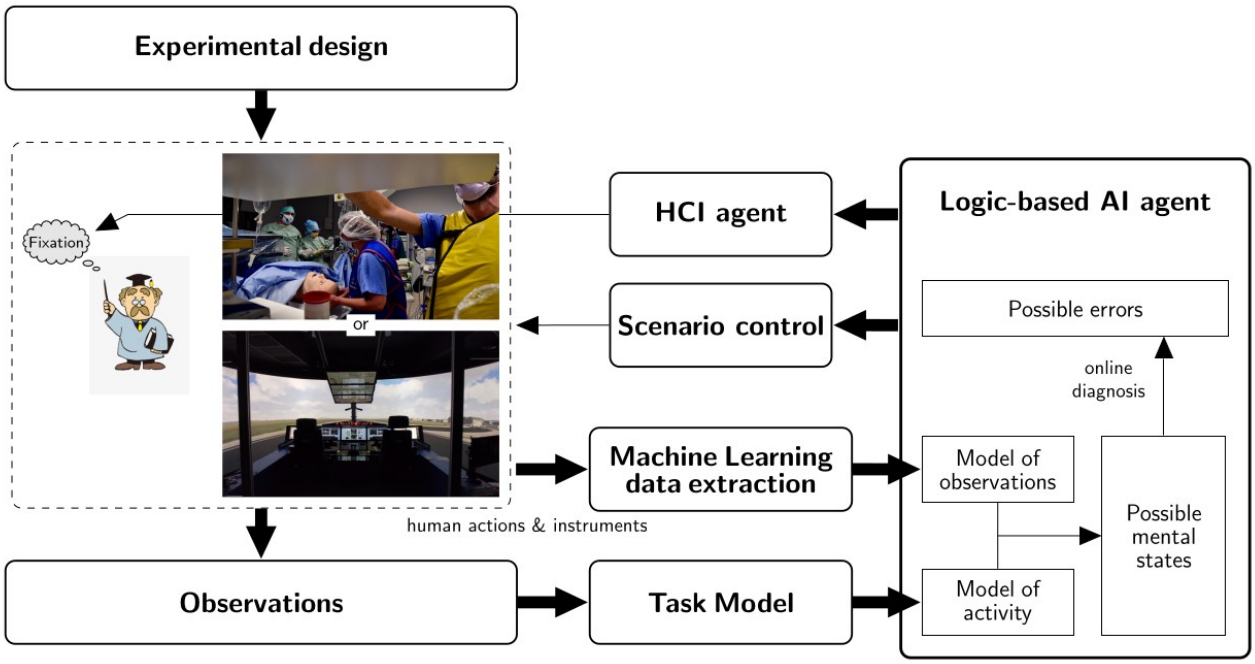

Figure 1 : Architecture générale du projet IDEFIX.
Le stage porte sur la brique « logic-based AI agent ».

Analyser les erreurs humaines à l'aide de la logique et de l'intelligence artificielle

Dans la plupart des accidents, les opérateurs se font une représentation erronée de la situation qui les conduit à prendre une mauvaise décision [1,2]. Les laboratoires LISN et LMF de l'université Paris-Saclay, deux laboratoires spécialisés dans la recherche en informatique, développent ensemble depuis près de 10 ans différents **modèles d'intelligence artificielle fondés sur la logique formelle** pour étudier ce type d'erreurs humaines. En particulier, dans la thèse de Valentin Fouillard [3], nous avons conçu, implémenté et validé un modèle informatique pour le diagnostic d'erreur humaine par reconstruction d'états mentaux cohérents. En partant d'une modélisation logique de la situation (dans laquelle sont représentées les observations et les actions des opérateurs), le système simule les croyances de l'opérateur et détermine les incohérences entre ces croyances et l'état du monde. Il calcule les corrections minimales à apporter pour retrouver un modèle cohérent (au sens de la logique (Tarski, 1946)) et en déduit les « états mentaux » des opérateurs qui pourraient expliquer leurs erreurs. Une modélisation en logique formelle des principaux biais cognitifs[1] [4,5] permet ensuite de filtrer les explications par type d'erreur [6,7].

L'objectif de ce stage est de poursuivre les travaux développés dans [3]. À partir de l'état de l'art des modèles de détection d'erreur (en particulier les travaux sur le monitoring d'activité humaine et les chroniques temporelles), proposer des solutions pour adapter l'approche précédente [3] au **monitoring temps-réel** d'opérateurs humains. L'un des objectifs de recherche du projet IDEFIX est en effet de passer du diagnostic à la prévention des erreurs.

Missions

Le stage se déroule en quatre étapes :

1. Vous devrez tout d'abord prendre en main les outils informatiques développés au LISN et au LMF pour le diagnostic de l'erreur humaine [3,6]. Le modèle logique utilise un format spécifique pour représenter les observations, les actions, les règles d'inférences et les croyances des opérateurs. Le moteur de raisonnement utilise le solveur SMT Z3 de Microsoft [8]. Il est implémenté en deux versions : C# et Java.
2. Vous devrez ensuite étudier un cas concret afin de représenter dans le modèle l'ensemble des données objectives nécessaires pour calculer des diagnostics possibles de l'erreur humaine. Le travail de modélisation se fera en collaboration avec des experts du Bureau d'Enquêtes et d'Analyse de l'aviation civile (BEA) ou avec des experts du domaine médical. Il vous amènera à étudier les limites du modèle logique. Vous analyserez ensuite les biais détectés par l'algorithme existant (**diagnostic** hors ligne).
3. Vous devrez faire un état de l'art des **méthodes en logique pour le monitoring d'activité** et la détection d'erreurs, en particulier l'analyse de chroniques temporelles [9,10].
4. Enfin vous proposerez une extension des algorithmes et du modèle logique utilisés pour le diagnostic (étape 1 et 2) afin de l'adapter à un contexte de surveillance en temps-réel (étape 3) dans le cadre du projet ANR IDEFIX.

Le stage se déroulera principalement au bâtiment 650 de l'université Paris-Saclay[2].

1 Les biais cognitifs sont des raisonnements rapides qui nous conduisent à prendre des décisions hâtives. Les erreurs de fixation font partie des biais cognitifs.

2 En cas de collaboration avec le BEA, il faudra prévoir quelques jours sur le site au Bourget pour travailler avec les enquêteurs et un engagement de confidentialité sera exigé pour l'accès aux données.

Contacts : Frédéric Boulanger frederic.boulanger@centralesupelec.fr et Nicolas Sabouret, nicolas.sabouret@centralesupelec.fr

Période de stage : entre février et septembre 2025 selon votre formation

Durée : 4 à 6 mois selon votre formation

Niveau requis : M2 mention informatique ou 3e année d'école d'ingénieur (option informatique)

Rémunération : env. 650 EUR/mois

Domaine : informatique, intelligence artificielle, logique formelle

Références bibliographiques :

[1] Dekker, S. (2006). *The field guide to understanding human error*. Cranfield University Press.

[2] Murata, A., Nakamura, T., and Karwowski, W. (2015). *Influence of cognitive biases in distorting decision making and leading to critical unfavorable incidents*. Safety, 1(1):44–58.

[3] Fouillard, V., Sabouret, N., Taha, S., and Boulanger, F. *An incremental diagnosis algorithm of human erroneous decision making*. In Proc. 2nd International Conference on Human and Artificial Rationalities (HAR), 2023.

[4] Tversky, A. and Kahneman, D. (1974). *Judgment under uncertainty : Heuristics and biases*. Science, 185(4157):1124–1131.

[5] Dimara, E., Franconeri, S., Plaisant, C., Bezerianos, A., and Dragicevic, P. (2020). *A task-based taxonomy of cognitive biases for information visualization*. IEEE Transactions on Visualization and Computer Graphics, 26(2):1413–1432.

[6] Fouillard, V., Sabouret, N., Taha, S., and Boulanger, F. *Catching cognitive biases in an erroneous decision making process*. In Proc. IEEE International Conference on Systems, Man and Cybernetics (SMC), 2021.

[7] Reason, J. (1990). *The contribution of latent human failures to the breakdown of complex systems*. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 327(1241):475–484.

[8] Moura, L. d. and Bjørner, N. (2008). *Z3: An efficient SMT solver*. Proc. International conference on Tools and Algorithms for the Construction and Analysis of Systems, pages 337–340. Springer.

[9] Guyet, T., & Besnard, P. (2023). *Chronicles: Formalization of a temporal model* (p. 121). Springer International Publishing.

[10] Dousson, C. (2002, July). Extending and unifying chronicle representation with event counters. In ECAI (Vol. 2, pp. 257-261).