Un modèle logique combinant la théorie de Dempster-Shafer et la théorie des possibilités pour la détection des erreurs de fixation

Gabrielle Porcher[1,2], Frédéric Boulanger[1], Nicolas Sabouret[2]

[1] Université Paris-Saclay, CNRS, ENS, CentraleSupélec, LMF, Gif-sur-Yvette, France
[2] Université Paris-Saclay, CNRS, LISN, Gif-sur-Yvette, France

Résumé

Les erreurs de fixation sont des erreurs d'origine humaine qui surviennent lorsqu'un individu se concentre de manière excessive sur une seule idée, une solution, une information ou une perspective, au point d'ignorer d'autres possibilités ou alternatives. Cet article présente un modèle fondé sur la logique formelle pour détecter des erreurs de fixation d'un opérateur humain dans une situation critique. Les informations communiquées à l'opérateur ainsi que ses actions sont représentées par des prédicats. Nous utilisons la théorie des fonctions de croyance pour calculer les états possibles du monde, puis nous utilisons le résultat dans le cadre possibiliste pour déterminer si les actions de l'opérateur sont cohérentes ou non pour alerter en temps réel l'opérateur sur un risque de fixation. Nous illustrons ce fonctionnement sur deux cas d'étude médicaux.

Mots-clés

Logique, Incertitude, Biais cognitifs, Théorie des possibilités, Théorie des fonctions de croyances

Abstract

Fixation errors are human-induced errors caused by an excessive focus on a single idea, solution, piece of information, or perspective, to the point of ignoring other possibilities. This paper presents a logic-based model for detecting fixation errors made by a human operator in a critical situation. The operator's actions and the information they receive are represented using predicates. We employ belief functions theory to compute the possible states of the world, and then utilize the possibilistic framework to determine whether the operator's actions are consistent with these states. This enables real-time alerts to be issued to the operator about a risk of fixation. We illustrate this approach with two medical case studies.

Keywords

Logics-Based Modeling, Uncertainty, Cognitive Biases, Possibility Theory, Theory of Belief Functions

1 Contexte et problématique

Le travail présenté dans cet article est réalisé dans le cadre du projet ANR IDEFIX (artificial Intelligence to DisEngage from FIXation) qui vise à décrire, étudier et prévenir les erreurs de fixations à l'aide d'un modèle d'intelligence artificielle. Dans des domaines comme la médecine ou l'aviation, dans lesquels la décision de l'opérateur met en jeu des vies humaines, il existe de nombreux outils ou méthodes pour aider l'opérateur dans sa décision. Les aides cognitives [6, 22] sont très utilisées : elles indiquent des conduites à suivre ou des éléments à vérifier mais elles n'aident pas toujours à identifier les bons diagnostics [3].

Les systèmes d'aide à la décision peuvent aussi être utilisés [17] mais ils présentent des risques de biais d'automatisation [13]. Dans le projet ANR IDEFIX, nous nous intéressons à une autre approche qui consiste à utiliser l'intelligence artificielle pour alerter les opérateurs lorsqu'ils se trouvent dans des situations où ils pourraient être victimes de biais de fixation. Le biais de fixation [9] est un cas particulier de biais cognitif [12] qui survient lorsqu'un individu se concentre de manière excessive sur une seule idée, une solution, une information ou une perspective, au point d'ignorer d'autres possibilités. Il est particulièrement critique en médecine ou en aviation et se retrouve sous différentes formes (tunnelisation, biais de confirmation, biais d'ancrage) [9]. Notre objectif n'est donc pas de dire à l'opérateur ce qu'il doit faire, comme dans un système d'aide à la décision, mais de l'alerter d'un risque potentiel.

L'un des enjeux dans le déploiement d'un tel outil est celui de l'explicabilité : un modèle « boîte noire » peut difficilement indiquer à l'opérateur les raisons de l'alerte. C'est pourquoi nous faisons le choix de nous appuyer sur un modèle en logique formelle qui manipule une représentation des informations et des actions de l'opérateur. Concrètement, dans le cadre du projet IDEFIX, les opérateurs sont confrontés à des scénarios (conçus par des experts métiers et des psychologues) favorisant l'apparition de biais de fixation. Le modèle informatique doit alors suivre en temps réel les évènements du scénario et alerter l'opérateur en cas de risque de fixation.

Nous espérons, à terme, montrer que l'IA permet de sortir plus rapidement de la fixation. Dans cet article, nous présentons une première version du modèle et du mécanisme d'alerte.

La construction d'un tel modèle soulève plusieurs problématiques :

1°) Pour créer des alertes pertinentes, il est nécessaire d'évaluer la possibilité d'une maladie (en médecine) ou d'une panne (en aviation) au cours du temps à partir des informations disponibles. Pour cela, il faut déterminer à quel point chaque information est associée à une maladie ou une panne, et comment ces associations se combinent quand plusieurs informations concernent une même maladie ou panne. Nous devons donc construire un **modèle logique capable d'effectuer des raisonnements de ce type**.

2°) Les informations transmises à l'opérateur sont parfois incomplètes ou incertaines. De même, les règles de raisonnement qui relient ces observations aux maladies ou pannes peuvent être ambiguës, ainsi que les règles qui décrivent les actions à effectuer pour les traiter. La raison en est qu'il n'est pas possible de capturer l'ensemble du réel dans des règles exactes et qu'il est donc nécessaire de manipuler des règles floues, avec de l'incertitude. Par exemple, une personne qui a la grippe a généralement de la fièvre, mais pas toujours. Un traitement est recommandé contre la COVID, sauf si le patient a des contre-indications, *et cetera*. Nous aurons donc besoin d'un **modèle logique capable de manipuler des incertitudes aussi bien au niveau des prédicats que des règles d'inférence**.

3°) Les experts humains utilisent des règles propres à leur métier pour réagir à des informations qu'ils reçoivent. Ces règles sont liées à des concepts situés à différents niveaux d'abstraction que les experts manipulent dans leur raisonnement. Par exemple lorsqu'un patient tousse, le médecin pense à une infection pulmonaire (concept abstrait), sans immédiatement décider entre une grippe ou une COVID (concepts plus concrets). Il faut donc que notre modèle métier supporte une modélisation hiérarchique des pannes et des maladies pour pouvoir effectuer des raisonnements sur des concepts de différents niveaux et les relier aux différentes actions. Mais surtout, ces actions peuvent nous renseigner sur le processus de décomposition des possibilités effectué par l'opérateur dans son diagnostic. Par exemple, demander un test bactérien permet d'isoler un sous-ensemble de maladies infectieuses sans pour autant savoir de quelle maladie précise il s'agit. Nous aurons donc besoin d'un **modèle métier qui permet de catégoriser des concepts de différents niveaux d'abstraction**.

Pour répondre à ces trois problématiques (et donc détecter un potentiel biais de fixation), nous proposons dans cet article un modèle logique capable de gérer l'incertitude et de suivre une exploration hiérarchique des pannes ou des maladies à l'aide d'un mécanisme d'inférence. Dans la prochaine section, nous présentons les travaux sur lesquels nous nous sommes appuyés pour construire ce modèle. Nous présentons dans la section 3 un premier cas d'étude qui servira d'exemple pour illustrer notre approche. Nous présentons le modèle lui-même dans la section 4, ainsi que le résultat obtenu sur notre cas d'étude. Dans la section 5 nous appliquons notre approche à un cas d'étude du projet IDEFIX. Enfin, nous discutons des limites et perspectives de ce modèle dans la section 7.

2 État de l'Art

Un de nos objectifs (problématique n° 1) est d'évaluer la possibilité des maladies ou pannes selon les évènements qui se présentent dans le scénario. Ces évènements peuvent être interprétés comme des symptômes déclenchés par les maladies et pannes elles-mêmes, ce qui se traduit par des règles logiques de la forme :

$$maladie/panne \rightarrow symptome$$

Dans ce cadre, retrouver l'origine du symptôme consiste à utiliser *l'abduction logique*. Des approches comme la logique abductive [14, 5] permettent de lister les causes possibles à partir des effets.
Dans nos travaux, nous avons besoin de combiner ces modèles abductifs avec un modèle de l'incertitude et avec un modèle de catégories hiérarchiques. Nous présentons dans les sections suivantes différents travaux allant dans ce sens.

2.1 L'incertitude

Pour modéliser le raisonnement de l'opérateur, nous avons besoin d'introduire de l'incertitude à la fois au niveau des prédicats et au niveau des règles. Il existe de nombreux travaux utilisant la logique pour le diagnostic médical, en particulier les systèmes experts tels que MYCIN [20] reposant sur des règles logiques et des facteurs de certitude, et les approches bayésiennes [15, 11]. Cependant, une difficulté est que nous n'avons pas connaissance des probabilités exactes associées à chaque fait et règle dans le contexte, ce qui rend difficile l'utilisation d'approches probabilistes, y compris des approches bayésiennes de l'abduction [16].

Logiques possibilistes

Un cadre logique permettant l'utilisation de poids non-probabilistes est le cadre de la logique possibiliste [8]. Elle vise à représenter et manipuler des connaissances incertaines qualitatives ou ordinales, sans recourir à une interprétation fréquentiste ou strictement probabiliste. Chaque formule logique est associée à un degré de nécessité (à quel point elle est nécessairement vraie) ou de possibilité (à quel point elle est possible d'après les informations disponibles), compris entre 0 et 1. Ces degrés ne sont pas interprétés comme des probabilités, mais comme des niveaux de compatibilité avec l'état du monde considéré. La combinaison des informations repose sur des opérateurs min et max, traduisant respectivement la conjonction prudente des contraintes, et l'agrégation de sources compatibles. Ce cadre a été utilisé avec succès pour représenter des raisonnements humains [18].
Cependant, dans nos travaux, nous faisons face à une difficulté supplémentaire : dans les scénarios sur lesquels nous travaillons, les observations sont nombreuses et peuvent donner des informations contradictoires. Dans le cadre possibiliste, la révision et la fusion d'informations se font avec les opérateurs min et max, ce qui avec nos données ne permet pas de conserver la finesse nécessaire au diagnostic. Si un symptôme faiblement associé à un diagnostic donné apparaît, les symptômes suivants, même très significatifs de

ce même diagnostic, seront absorbés sans pouvoir faire augmenter le degré de possibilité du diagnostic car l'algorithme prendra le minimum des valeurs de possibilité.

Exemple

On cherche à maximiser la mesure de possibilité de l'hypothèse H, qui peut être soit $Pneumonie$, soit $Covid$. On observe deux symptômes : des nausées (représentées par le prédicat $nausee$) puis un résultat de scanner caractéristique d'une pneumonie (que nous noterons $scan\_p$). La nausée est très peu associée aux deux hypothèses (0.1) tandis que le résultat au scanner est très associé à la pneumonie (0.9) et très peu à la COVID (0.1). La mesure de possibilité conjointe est donnée par :

$$\pi(nausee, scan\_p|H) = \min(\pi(nausee|H), \pi(scan\_p|H))$$

Si $H = Pneumonie$:
$\pi(nausee, scan\_p|Pneumonie) = \min(0.1, 0.9) = 0.1$

Si $H = Covid$:
$\pi(nausee, scan\_p|Covid) = \min(0.1, 0.1) = 0.1$

Dans cet exemple, on dit alors que les deux hypothèses $Pneumonie$ et $Covid$ expliquent les symptômes avec la même mesure de possibilité (0.1). Pourtant, on voudrait pouvoir associer l'hypothèse $Pneumonie$ à une valeur plus élevée, étant donné qu'elle est bien plus fortement associée au résultat de scanner obtenu.

Théorie des fonctions de croyance

Une autre approche proposée dans la littérature est la théorie des fonctions de croyance, introduite par Dempster [7] et formalisée par Shafer [19], qui introduit la notion de « probabilités incertaines » : la probabilité associée à chaque fait ou règle peut être comprise dans un intervalle (au lieu d'une valeur exacte). Elle permet ainsi de calculer la probabilité associée à chaque panne ou maladie à partir d'informations et de règles imprécises ou incertaines. En particulier, dans un cadre abductif, elle permet de raisonner sur des situations dans lesquelles un même symptôme peut être obtenu de manière incertaine à partir de pannes ou maladies différentes. On définit pour cela une *fonction de masse* qui attribue une valeur à chaque sous-ensemble de l'espace des hypothèses. La masse d'un ensemble représente la proportion de faits qui supportent l'hypothèse que l'état du monde soit dans cet ensemble et non dans un des sous-ensembles possibles. Elle représente donc à la fois le degré de certitude que l'état soit dans un ensemble, et l'absence d'information permettant de préciser l'état exact du monde. La règle de combinaison de Dempster permet alors d'agréger deux fonctions de masse indépendantes m_1 et m_2 définies sur le même ensemble d'hypothèses :

$$m = m_1 \oplus m_2$$

Nous illustrerons ce calcul dans la section 4. Nous montrerons qu'il n'y a alors pas d'effet d'absorption lié aux opérateurs min et max.

Remarque

La théorie des fonctions de croyances suppose un cadre de discernement (c'est-à-dire l'ensemble des abductibles exprimés) qui est fermé. En d'autres termes, l'ensemble des pannes ou maladies possibles doit être exhaustif. C'est le choix que nous avons fait dans notre modèle même si cette hypothèse pose plusieurs problèmes que nous discuterons dans la section 7.

2.2 La catégorisation

Le deuxième problème auquel nous faisons face est que le raisonnement des experts ne se situe pas directement au niveau des maladies mais utilise des catégories de plus haut niveau. Ce problème a été bien étudié par la communauté de l'ingénierie des connaissances, par exemple avec les logiques de description [1]. Ces modèles permettent de raisonner avec une représentation hiérarchique de la connaissance. Par exemple, on peut décrire qu'une atteinte pulmonaire provoque de la toux et en déduire que la grippe, qui est une sous-classe des atteintes pulmonaires, provoque ce symptôme. Dans notre cas, les actions menées par les opérateurs lors de leur exploration des possibilités peuvent être associées à des catégories de maladies ou pannes de plus haut niveau ; les représenter permet donc de suivre leur raisonnement. Par exemple, une radiographie thoracique peut être réalisée lorsqu'une atteinte pulmonaire est soupçonnée, mais cela ne permet pas de savoir si l'opérateur envisage une atteinte en particulier, comme une pneumonie.
Dans cet article, nous allons utiliser ce type de représentation et combiner cette catégorisation avec la théorie des fonctions de croyances pour mesurer la nécessité de la réalisation ou non des actions associées aux catégories.

3 Cas d'étude

Le modèle présenté dans la section suivante a été testé sur un scénario tiré d'un exemple réel dans lequel un médecin donne un traitement inadapté en raison d'un biais de fixation de type « tunnelisation » (c'est-à-dire une attention focalisée sur un symptôme ou une maladie particulière) décrit dans [2]. Nous utiliserons ce scénario pour illustrer les différents composants de notre modèle. L'implémentation de ce modèle est disponible sur GitHub [1].

Le scénario se déroule ainsi : un patient se présente avec un ensemble de symptômes (toux, fièvre, difficultés respiratoires). Le médecin en charge fait alors une série d'actions épistémiques (examens, tests, scanner) et reçoit des résultats. Il explore le diagnostic le plus probable, une pneumonie bactérienne typique, et donne le traitement associé. N'observant pas d'amélioration de l'état du patient, ce qui suggère un mauvais diagnostic, le médecin se tourne ensuite vers un diagnostic très peu probable, la COVID, vraisemblablement en raison du contexte de pandémie au moment des faits, et cela malgré des informations orientant vers une pneumonie atypique. Il donne donc le mauvais traitement alors qu'il disposait des bonnes informa-

1. https://github.com/GabriellePorcher/RJCIA_2026

tions (test COVID négatif, scanner suggérant fortement une pneumonie).

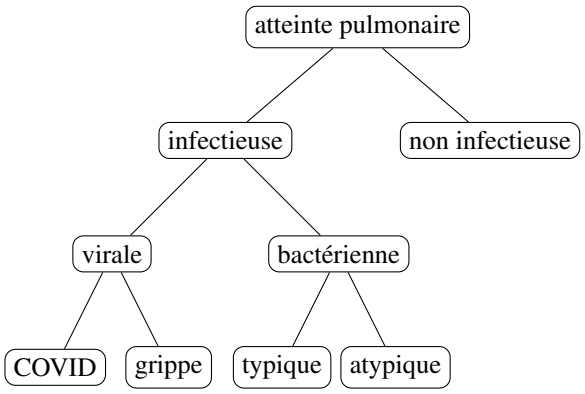

FIGURE 1 – Hiérarchie des atteintes pulmonaires

La hiérarchie des maladies que nous considérons est représentée sur la figure 1. Comme discuté dans la section précédente, il s'agit d'un monde fermé (alors que cet arbre est forcément très incomplet). Cette hiérarchie nous permet d'associer des actions à différentes catégories de maladies.

4 Modèle proposé

4.1 Principe général

Nous proposons de définir la fixation comme un écart entre la possibilité d'une maladie et à quel point l'opérateur semble l'envisager. Ainsi, un opérateur qui semble se concentrer sur une possibilité considérée comme peu probable par le modèle sera alerté d'un potentiel biais de fixation. L'algorithme évalue d'abord les pannes ou maladies possibles à partir des observations pour déterminer celles qui sont les plus probables. Il évalue ensuite les potentielles erreurs de l'opérateur en observant ses actions.
Notre proposition se divise donc en deux grandes parties :

— La détermination de la probabilité de la maladie ou panne, à partir des observations ;
— La détermination de la nécessité d'exécuter une action d'après des règles métier.

Selon le scénario, l'opérateur peut avoir à diagnostiquer une panne, une anomalie... Dans cette partie, de part le choix de notre cas d'étude, nous parlerons de suivi des maladies pour décrire les pistes réelles pouvant être explorées par les opérateurs.

4.2 Associations observations → maladies

Cette première partie de notre modèle s'appuie sur la théorie des fonctions de croyances de Dempster-Shafer [19] et la définition de fonctions de masses.

Considérons :

- $\boldsymbol{\Omega} = \{d_1, \ldots, d_n\}$ le cadre de discernement, correspondant à l'ensemble fini des maladies possibles ;
- $\boldsymbol{O} = \{o_1, \ldots, o_k\}$ l'ensemble des observations possibles (symptômes, signes, résultats d'examens) ;
- $\boldsymbol{2^{\Omega}}$ l'ensemble des sous-ensembles de Ω.

Pour chaque observation $o \in O$, nous devons définir une fonction de masse sur l'ensemble des diagnostics possibles :

$$m_o : 2^{\Omega} \to [0, 1]$$

telle que :

$$\sum_{S \subseteq \Omega} m_o(S) = 1 \quad \text{et} \quad m_o(\varnothing) = 0.$$

Interprétation :

- $\boldsymbol{m_o(\{d_i\})}$ représente le poids directement attribué à la maladie d_i par l'observation o ;
- $\boldsymbol{m_o(S)}$ avec $|S| > 1$, représente le poids associé aux éléments de S, sans discrimination (l'observation soutient un ensemble de maladies sans distinction) ;
- $\boldsymbol{m_o(\Omega)}$ modélise l'incertitude globale associée à l'observation.

Illustration sur le cas d'étude

Dans notre exemple, nous considérons 4 maladies (pneumonie typique, pneumonie atypique, grippe et COVID) :

$$\Omega = \{pneumo\_t, pneumo\_a, grippe, covid\}$$

et un ensemble d'observations correspondant aux symptômes, aux résultats d'examens cliniques ou à l'état général du patient :

$$O = \{toux, fievre, sat, rales, scanb, ac\}$$

où $toux$ et $fievre$ correspondent aux symptômes décrits par le patient à son arrivée, sat représente une saturation en oxygène normale mesurée par l'oxymètre, $rales$ représente la présence de râles crépitants à l'auscultation pulmonaire, $scanb$ représente la présence d'une structure en arbre en bourgeon lors du scanner thoracique et ac représente une amélioration générale de l'état du patient.
Il y a bien sûr d'autres observations (symptômes, examens ou description de l'état du patient) dans le scénario complet mais nous nous limiterons à cette liste pour illustrer notre modèle dans cet article.

Les valeurs attribuées aux fonctions de masse de chaque observation pour les maladies sont définies arbitrairement (elles nous permettent simplement ici d'illustrer les propriétés du modèle, pas de valider un résultat médical).
Voici un exemple de fonction de masse : l'observation d'une structure en arbre en bourgeons au scanner thoracique est un signe radiologique indiquant plutôt une pneumonie. Nous pouvons le représenter par :

$$m_{scanb} = \begin{cases} 0.8 & \to pneumo, \\ 0.05 & \to grippe, \\ 0.05 & \to covid, \\ 0.1 & \to \Omega. \end{cases}$$

avec $pneumo = \{pneumo\_t, pneumo\_a\}$ le sous-ensemble regroupant les deux pneumonies. L'observation $scanb$ supporte donc fortement l'hypothèse « pneumonie », sans permettre de discriminer les deux types de pneumonie. Remarquons que tous les sous-ensembles de Ω ne sont pas associés à une valeur. La masse des ensembles non-mentionnés est nulle. La masse d'Ω représente l'incertitude globale sur le diagnostic.

4.3 Calcul du score à partir des observations

Considérons $A \subseteq \Omega$ un ensemble de maladies et deux observations o_1 et o_2 dont on souhaite calculer la combinaison des fonctions de masse $(m_{o_1} \oplus m_{o_2})(A)$. Dans la théorie de Dempster-Shafer, le calcul de $\oplus$ se fait de la manière suivante :

$$(m_{o_1} \oplus m_{o_2})(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_{o_1}(B)\, m_{o_2}(C)$$

où :

$$K = \sum_{B \cap C = \varnothing} m_{o_1}(B)\, m_{o_2}(C)$$

représente le degré de conflit entre les deux sources d'information, mesurant la contradiction entre ces sources.
Cet opérateur combine donc les fonctions de masse de tous les sous-ensembles qui soutiennent l'hypothèse A pour les deux observations.
Lorsque plusieurs observations sont faites dans un scénario, nous notons $m(A)$ le score final associé au diagnostic A après combinaison des fonctions de masses de toutes les observations (A peut aussi bien être une maladie dans Ω, ou une catégorie de maladies dans 2^Ω) :

$$m(A) = (m_{o_1} \oplus \cdots \oplus m_{o_k})(A)$$

Illustration sur le cas d'étude

Pour illustrer cette combinaison des fonctions de masse, considérons deux observations : une saturation en oxygène normale (sat), plutôt indicatrice de grippe ou de COVID, et la présence de râles crépitants à l'auscultation pulmonaire ($rales$), plutôt indicateurs d'une pneumonie, sans que cela permette de distinguer le type de bactérie responsable. Les valeurs des fonctions de masse pour ces deux observations sont décrites dans le tableau suivant :

A	$m_{sat}(A)$	$m_{rales}(A)$
$\{pneumo\}$	0.2	0.60
$\{grippe\}$	0.30	0.10
$\{covid\}$	0.40	0.20
Ω	0.10	0.10

Nous calculons alors la valeur du conflit K (ce qui est assez simple dans notre cas puisque tous les éléments sauf Ω sont disjoints) :

$$\begin{aligned} K &= 0.2 \times (0.1 + 0.2) \\ &\quad + 0.3 \times (0.6 + 0.2) \\ &\quad + 0.4 \times (0.6 + 0.1) \\ &= 0.58. \end{aligned}$$

Les masses combinées sont alors déterminées comme suit (avec $1 - K = 0.42$) :

$$\begin{aligned} m(\{pneumo\}) = &\\ &\tfrac{1}{0.42}(m_{sat}(\{pneumo\})) \cdot m_{rales}(\{pneumo\}) \\ &+ m_{sat}(\{pneumo\}) \cdot m_{rales}(\{\Omega\}) \\ &+ m_{sat}(\{\Omega\}) \cdot m_{rales}(\{pneumo\})) \\ &\approx 0.476 \end{aligned}$$

$$\begin{aligned} m(\{grippe\}) = &\\ &\tfrac{1}{0.42}(m_{sat}(\{grippe\}) \cdot m_{rales}(\{grippe\}) \\ &+ m_{sat}(\{grippe\}) \cdot m_{rales}(\{\Omega\}) \\ &+ m_{sat}(\{\Omega\}) \cdot m_{rales}(\{grippe\})) \\ &\approx 0.167 \end{aligned}$$

$$\begin{aligned} m(\{covid\}) = &\\ &\tfrac{1}{0.42}(m_{sat}(\{covid\}) \cdot m_{rales}(\{covid\}) \\ &+ m_{sat}(\{covid\}) \cdot m_{rales}(\{\Omega\}) \\ &+ m_{sat}(\{\Omega\}) \cdot m_{rales}(\{covid\})) \\ &\approx 0.334 \end{aligned}$$

$$\begin{aligned} m(\Omega) = &\\ &\tfrac{1}{0.42}(m_{sat}(\Omega) \cdot m_{rales}(\Omega)) \\ &\approx 0.024 \end{aligned}$$

Remarquons que la combinaison des masses par $\oplus$ permet bien d'obtenir des valeurs reflétant les informations données par les deux symptômes, où les poids peuvent être renforcés et affaiblis : ce type de résultat permet donc de déterminer à quel point les maladies sont soutenues par toutes les observations, et donc la pertinence qu'aurait le médecin à les explorer (actions épistémiques) ou à les traiter (actions curatives).

4.4 Modélisation des règles liées aux actions

La deuxième partie de notre modèle consiste à déterminer la nécessité d'une action. Cela permet d'alerter l'opérateur (*i.e.* le médecin dans notre exemple) d'un possible biais de fixation s'il ne la fait pas. Pour cela, nous allons nous appuyer sur la représentation hiérarchique des maladies.

Hiérarchie des maladies

Nous notons $\mathcal{H}$ la hiérarchie structurant les concepts de différents niveaux que manipulent les opérateurs dans un contexte donné (et dont un exemple est donné sur la figure 1). Ces concepts sont soit des éléments du cadre de discernement Ω, soit des concepts plus abstraits, par exemple $infection\_bacterienne$ qui regroupe les deux cas de pneumonie présents dans Ω.
Notons que la hiérarchie $\mathcal{H}$ peut aussi inclure des nœuds ne correspondant à aucune maladie spécifique de notre étude, tels que la catégorie « non infectieuse » sur la figure 1. Bien qu'aucun élément de Ω n'appartienne à cette catégorie, sa

représentation reste nécessaire dans notre modèle : elle permet d'associer des actions spécifiques d'un médecin en présence de ce type de pathologie. Les éléments de Ω sont nécessairement des feuilles de $\mathcal{H}$ car ce sont les diagnostics les plus concrets auxquels nous nous intéressons.

Le savoir-faire de l'opérateur détermine les actions à entreprendre en fonction de la catégorie de maladie suspectée. Nous représentons ces connaissances de l'expert sous forme de règles indiquant la nécessité $\alpha \in [0, 1]$ de réaliser une action a en présence d'un diagnostic $C \in \mathcal{H}$:

$$C \overset{\alpha}{\Rightarrow} a$$

Le degré de nécessité α représente la nécessité de l'action lorsque toutes les prémisses sont pleinement satisfaites.
Par exemple, en présence certaine d'une atteinte pulmonaire infectieuse d'origine bactérienne, il est nécessaire à 90% de prescrire des antibiotiques :

$$bacterienne \overset{0.9}{\Longrightarrow} antibio$$

Pour déterminer la nécessité d'une action, il nous faut combiner la nécessité d'appliquer une règle et la certitude que nous avons sur un diagnostic. La théorie de Dempster-Shafer qui a été pertinente pour fusionner les informations issues des observations (comme nous l'avons montré dans la partie précédente), ne s'applique pas ici. En effet, il ne s'agit pas de combiner différentes sources pour en déduire des faits, mais de mettre en regard la confiance dans un diagnostic et la nécessité de faire une action. Nous nous plaçons donc dans un cadre possibiliste.
Dans ce contexte, nous interprétons les fonctions de masse précédemment obtenues comme des degrés de nécessité au sens possibiliste, c'est-à-dire la nécessité que le patient soit atteint d'une maladie donnée.
Ainsi, pour tout concept $C \in \mathcal{H}$, nous définissons son degré de nécessité $N(C)$ comme suit :

— Si $C \in \Omega$, le degré de nécessité de la maladie est la valeur donnée par la combinaison des fonctions de masse pour toutes les observations :

$$N(C \in \Omega) = m(C)$$

— Sinon, le degré de nécessité d'un concept de plus haut niveau est défini comme la somme de sa masse et des degrés de nécessité de ses fils :

$$N(C \notin \Omega) = m(C) + \sum_{D \; is\text{-}a \; C} N(D)$$

avec $is\text{-}a$ la relation entre deux sommets de $\mathcal{H}$: $D \; is\text{-}a \; C$ si D est fils direct de C dans l'arbre $\mathcal{H}$.

Notons que la nécessité d'un concept qui ne généralise aucun élément de Ω (comme « non infectieuse » dans notre exemple) sera toujours nulle. Cela correspond au fait que, dans le scénario considéré, cette catégorie de maladies est extrêmement peu probable et aurait dû être écartée par l'opérateur.

Nécessité d'une action

Pour une règle donnée, en suivant le cadre de la logique possibiliste présenté à la section 2, nous pouvons calculer le degré de nécessité associé à l'action a comme la combinaison, avec l'opérateur $\min$, du degré de nécessité de C et de celui de la règle :

$$N(a) = \min(\alpha, \, N(C))$$

Dans le cas général, lorsque plusieurs règles métier peuvent conduire à la même action, nous combinons ces règles avec l'opérateur $\max$:

$$N(a) = \max_{C \overset{\alpha}{\Rightarrow} a}(\min(\alpha, \, N(C)))$$

Ainsi, le degré de nécessité d'une action correspond au maximum, pour toutes les règles qui s'appliquent, du minimum entre :

— le degré de nécessité de la prémisse,
— le degré de nécessité de la règle logique associée.

L'utilisation du $\max$ pour combiner les nécessités induites par plusieurs règles, plutôt que d'utiliser Dempster-Shafer, se justifie par le fait que l'on combine des nécessités et pas des masses : à partir du moment où une action est nécessaire, cette nécessité ne peut pas être amoindrie par le fait qu'une autre règle la rend aussi nécessaire, mais à un degré moindre.

Illustration sur le cas d'étude

Dans notre scénario, supposons que nous avons la règle suivante :

$$covid \overset{0.9}{\Longrightarrow} traitement\_covid$$

Si $m(\{covid\}) = 0.1$ alors le degré de nécessité associé à l'action $traitement\_covid$ est $\min(0.1, 0.9) = 0.1$.

Seuils d'alerte

Pour déterminer quand alerter l'opérateur qui n'effectue pas une action nécessaire, nous définissons trois valeurs :

— un seuil d'alerte σ_{min} qui représente le degré de nécessité en-dessous duquel une action ne devrait jamais être faite ;
— un seuil d'alerte σ_{max} qui représente le degré de nécessité à partir duquel une action devrait absolument être faite ;
— une durée δ qui représente le délai maximum avant une alerte pour les actions non effectuées.

Notre modèle déclenche une alerte dans deux cas :

— l'opérateur fait l'action a alors que $N(a) < \sigma_{min}$,
— il existe une action a telle que $N(a) > \sigma_{max}$ depuis au moins δ, et qui n'a pas été effectuée.

La durée δ permet de prendre en compte le temps nécessaire à l'opérateur pour effectuer l'action, et d'éviter les alarmes intempestives.
Dans ces deux situations, nous considérons qu'il y a un risque d'erreur de fixation de la part de l'opérateur.

4.5 Résultats sur le cas d'étude

Nous avons implémenté le scénario de notre cas d'étude (voir section 3) avec des fonctions de masses définies comme illustré dans la sous-section 4.3. Nous avons choisi $\sigma_{min} = 0.2$ comme seuil d'alerte pour les actions effectuées et $\sigma_{max} = 1$. Nous ne déclenchons donc pas d'alerte pour les actions non-effectuées car nous ne considérons dans ce scénario que les cas de fixation entraînant une action non justifiée, en raison de l'absence d'information sur les autres actions possibles [2].
La figure 2 illustre l'évolution des masses des quatre maladies, de l'ensemble Ω (ignorance totale) ainsi que les valeurs du conflit K. Nous voyons que la COVID est plus probable au début du scénario et diminue assez vite (après les premiers examens) alors que la masse associée à la pneumonie atypique augmente rapidement et domine les masses des autres maladies. La valeur de K qui reste élevée montre un conflit significatif pendant la majeure partie du scénario, jusqu'à ce que la bactérie responsable de la pneumonie atypique soit découverte par un test.

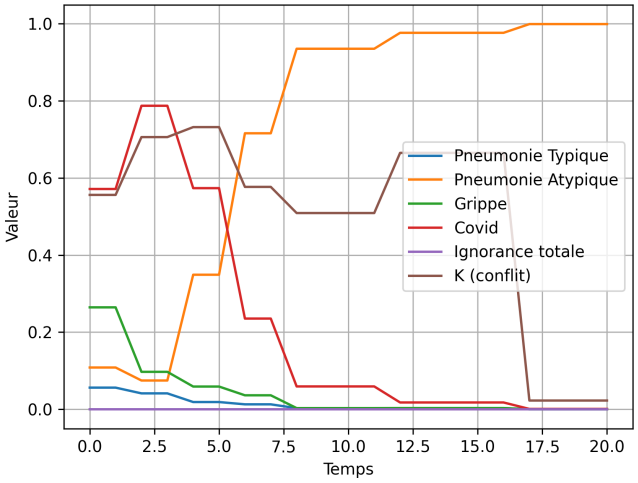

FIGURE 2 – Masses et conflit cas n° 1

Au cours du scénario, le médecin effectue plusieurs actions. Notre modèle donne une alerte pour quatre d'entre elles :

— La première à $t = 9$ lorsqu'il donne un traitement permettant de soigner les pneumonies typiques. À cette date, la masse associée à ce type de pneumonie est particulièrement faible.
Dans la réalité, il n'est pas anormal d'explorer cette piste en premier. En l'absence d'information complémentaire, ce traitement va permettre au médecin soit de guérir le patient, soit d'écarter la maladie la plus probable en l'absence d'amélioration.

— La seconde alerte à $t = 11$ concerne un test COVID : au lieu de persister dans la catégorie des infections bactériennes, qui est l'explication la plus plausible, le médecin décide de tester la COVID qui est peu probable au vu des informations.
Dans la réalité, le contexte général de la pandémie en 2020 explique le choix de cette action épistémique, peu invasive.

— Malgré un test COVID négatif, le médecin décide de donner un traitement anti COVID ($t = 13$) et de réaliser un second test COVID ($t = 14$).
Il s'agit là manifestement d'erreurs de fixation. Notre modèle déclenche des alertes car la valeur attribuée à COVID continue de décliner.

Ce scénario simple montre deux choses : premièrement, que toutes les situations d'alerte ne correspondent pas nécessairement à des erreurs de fixation ($t = 9$ et $t = 11$) : le médecin peut avoir de bonnes raisons de faire une action qui ne semble pas nécessaire. Deuxièmement, malgré la simplicité des règles métier utilisées, notre modèle est capable de repérer une situation de fixation. Nous pouvons penser que si le médecin avait été alerté, il n'aurait peut-être pas fait l'erreur de prescrire un traitement anti COVID. C'est justement ce que nous voulons étudier dans le projet IDEFIX.

5 Un scénario du projet IDEFIX

Dans le cadre du projet ANR IDEFIX, des scénarios inspirés de situations réellement vécues à l'hôpital ont été construits dans un but de formation des élèves médecins. Ils sont conçus spécialement pour conduire à une fixation. Dans un de ces scénarios, le patient (qui est un mannequin de simulation médicale dans nos expériences) présente des troubles neurologiques et des symptômes infectieux, ce qui conduit à un premier diagnostic de méningite (cohérent avec ces symptômes). Des informations ultérieures viennent contredire ce diagnostic initial (le patient souffre en réalité de deux pathologies : une grippe, expliquant les symptômes infectieux, et un AVC, expliquant les troubles neurologiques). Une trentaine d'élèves de l'école de médecine de Lyon ont été confrontés à ce scénario et enregistrés (vidéo, audio et données du simulateur médical utilisé). Les informations nécessaires à l'application de notre modèle ont été extraites d'un de ces enregistrements afin de nous permettre d'évaluer son potentiel sur un cas réel.

Modélisation des hypothèses multiples

Ce scénario présente une problématique particulière : le patient présente conjointement deux pathologies. Le cadre de discernement (l'ensemble des hypothèses envisageables dans la théorie des fonctions de croyances) est ici constitué de la méningite, la grippe, l'AVC, et l'occurrence simultanée d'une grippe et d'un AVC : les hypothèses du cadre de discernement devant être mutuellement exclusives, nous gérons la présence de plusieurs pathologies simultanées en les ajoutant explicitement au cadre de discernement. Ainsi, $\{grippe\}$ correspond à un patient atteint de grippe seulement, tandis que $\{grippe\_avc\}$ correspond à un patient atteint de grippe et faisant un AVC.
Pour gérer l'attribution des masses dans ce cadre, lorsqu'un symptôme est lié à deux maladies du cadre de discernement qui peuvent être simultanées, on attribue la masse au sous-ensemble comprenant tous les éléments du cadre où la maladie figure. Par exemple, si le test de la grippe revient positif, on attribue la masse au sous-ensemble $\{grippe, grippe\_avc\}$ indiquant qu'une de ces deux hypo-

thèses est vraie (grippe seule ou grippe avec AVC) sans pouvoir déterminer laquelle d'après le résultat de test.
De la même manière, dans la modélisation de la nécessité des actions, si la grippe entraîne une action, on écrira que $grippe \vee grippe\_avc$ implique l'action : par exemple, $grippe \vee grippe\_avc \overset{\alpha}{\Longrightarrow} tamiflu$ signifie que si le minimum de α et de la masse associée à la grippe seule ou à la grippe avec AVC dépasse un seuil donné, le médecin devrait donner au patient le médicament antiviral Tamiflu, indiqué contre la grippe.

Entrées du modèle : observations du médecin

Nous distinguons les observations de l'opérateur, qui nous renseignent sur l'état du monde, et ses actions, qui nous renseignent sur les hypothèses qu'il envisage. Toutes les observations connues du participant sont utilisées en entrée du modèle de diagnostic (Dempster-Shafer), qu'elles soient présentes dans le briefing pré-intervention ou qu'elles surviennent lors du déroulement du scénario. Sont donc pris en compte des facteurs de risques (âge, fumeur, hypertension), des constantes vitales (fréquence cardiaque, pression artérielle, saturation en oxygène), des symptômes apparaissant spontanément (état des pupilles, niveau de conscience), et des résultats de tests ou d'actions réalisées (résultat de ponction lombaire, réactions à un traitement).

Entrées du modèle : actions du médecin

Les actions de l'élève médecin comprennent des actions qu'il effectue directement et des actions qu'il délègue à l'infirmier anesthésiste. Dans notre modèle, l'action « examen des pupilles » représente donc aussi bien l'examen des pupilles par le médecin que le fait qu'il demande à l'infirmier de les examiner. Nous distinguons également les actions de prise d'information, dites épistémiques (comme l'examen des pupilles), des actions de traitement, dites pragmatiques (par exemple administrer un antibiotique). Parmi ces dernières, certaines actions ne servent qu'à stabiliser l'état du patient et ne sont pas spécifiques à une pathologie particulière. Nous avons décidé de ne pas les représenter dans le modèle, car elles ne donnent pas lieu à des biais de fixation. Nous n'avons pas non plus représenté les actions de verbalisation des hypothèses (« je crois qu'il a la grippe ») mais nous verrons plus tard qu'elles pourraient être utiles pour valider notre modèle.
Comme dans le premier cas d'étude, les pathologies considérées sont regroupées en catégories dans la hiérarchie représentée sur la figure 3. Ces catégories sont utilisées comme prémisses dans les règles de nécessité des actions.

Résultats

L'évolution des masses des différentes pathologies au cours du scénario est illustrée sur la figure 4. À $t = 0$, les masses sont assez précises étant donné que les nombreuses informations données dans le briefing ont déjà été enregistrées. Si pendant la première moitié du scénario, l'hypothèse de la méningite est favorisée, elle diminue nettement à $t = 16$ au profit de l'hypothèse de la grippe et de l'AVC conjoints, lorsqu'on apprend que le patient est positif à la grippe. Cette

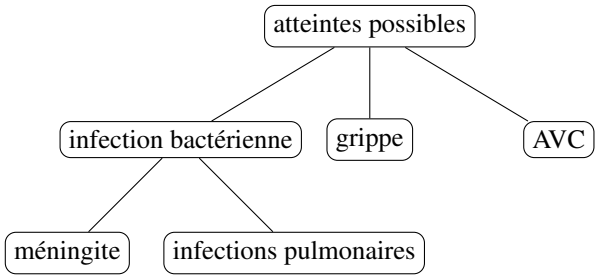

FIGURE 3 – Hiérarchie de pathologies du scénario

tendance se poursuit à $t = 28$ lorsque le patient présente une mydriase, un état de dilatation de la pupille caractéristique de l'AVC. Nous pouvons aussi remarquer que la masse de l'AVC seul reste à zéro car, dès le début de la simulation, il y a des indices clairs de maladie infectieuse, qui rend impossible ce singleton. De même, la masse de la grippe seule reste assez basse en raison des symptômes qui favorisent les sur-ensembles incluant aussi des affections neurologiques.

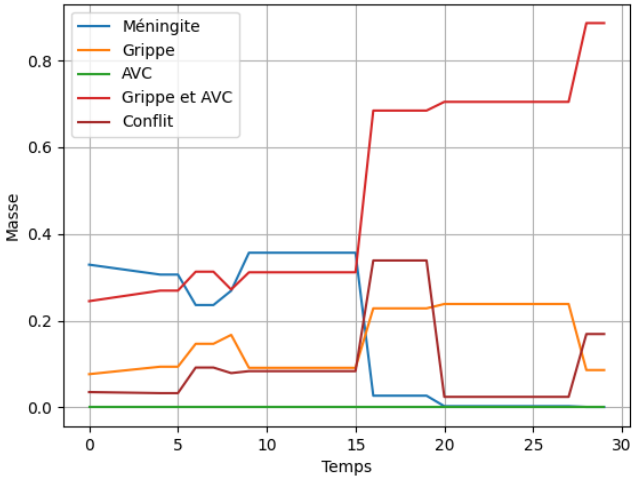

FIGURE 4 – Masses et conflit cas n° 2

Dans ce scénario, nous obtenons six alertes (avec les seuils $\sigma_{max} = 0.7$ et $\sigma_{min} = 0.3$) en réaction aux actions de l'élève médecin. À partir de $t = 16$, alors que la masse de l'ensemble avec grippe et AVC passe au dessus de σ_{max}, nous recevons une première alerte indiquant que l'action « réaliser un scanner avec injection » (scanner permettant de détecter l'AVC) doit être réalisé, étant donné que le degré de nécessité associé à cette règle est également supérieur au seuil fixé. Il s'agit de la seule action spécifique à l'AVC n'ayant pas encore été réalisée par l'opérateur dans ce scénario. Dans la réalité, il sera intéressant de notifier ce type d'action, qui pourrait permettre à l'opérateur d'envisager à nouveau la piste de l'AVC.
Les cinq alertes déclenchées ensuite correspondent à des actions spécifiques soit à la méningite, soit à des infections respiratoires, qui n'auraient pas dû être réalisées d'après le modèle, telle qu'un scanner thoracique. En réalité, bien que

les maladies détectables par ce test soient associées à des masses très faibles, réaliser ce type de test peut toujours être intéressant et surtout peu coûteux pour le patient s'il est stable.

En conclusion, le modèle propose des alertes cohérentes sur ce scénario en jugeant correctement les hypothèses les plus probables, et serait capable de notifier un utilisateur à la fois pour suggérer des actions et mettre en en avant des actions réalisées mais non-nécessaires, pouvant être causées par un biais de fixation. Il sera intéressant d'appliquer ce modèle avec les mêmes paramètres à d'autres participants à l'étude sur ce même scénario pour observer si les alertes restent pertinentes. Les prochaines étapes du projet IDEFIX vont consister à voir si les indications de notre modèle permettent effectivement au médecin, sur ces mêmes scénarios, de sortir plus tôt de la fixation.

6 Discussion

L'approche permettant l'évaluation des probabilités par la théorie des fonctions de croyances semble montrer de bons résultats sur le cas d'étude et le scénario du projet que nous avons utilisés, mais elle présente aussi plusieurs limites. En particulier, lorsque la valeur de K est trop élevée, Zadeh montre dans [23] qu'il faut conserver une certaine prudence dans l'interprétation des résultats. En effet, lorsque les sources sont contradictoires, la normalisation amplifie les masses des sous-ensembles non vides. Dans notre premier exemple, en sortie du modèle, une seule maladie domine très largement les autres, alors que dans la réalité plusieurs pistes sont probables en particulier en début de scénario alors que peu d'informations sont encore disponibles. De plus, dans notre modèle, nous avons travaillé avec un cadre de discernement fermé (considérant que toutes les hypothèses sont représentées dans Ω) et avec des hypothèses mutuellement exclusives, comme le propose Shafer [19]. Pourtant, travailler avec un cadre ouvert et pouvoir représenter le fait que plusieurs hypothèses peuvent être vraies en même temps serait plus proche de la réalité à laquelle sont confrontés les opérateurs.

Enfin, l'attribution des masses est aussi une problématique importante pour l'utilisation de notre modèle. Elle nécessite une expertise métier dans un contexte spécifique à chaque situation. Dans notre exemple, le choix des masses permet de produire trois bonnes alertes mais il cause aussi une alerte qu'il aurait été préférable d'éviter, alors que l'opérateur explore la possibilité d'une pneumonie typique.

La section suivante propose des pistes de résolution pour les problématiques identifiées ici.

7 Conclusion et perspectives

Ce travail propose un modèle formel pour la détection en temps réel du biais de fixation chez les agents humains dans des contextes critiques, en combinant deux approches complémentaires : une modélisation de la réalité via la théorie des fonctions de croyances (Dempster-Shafer) pour estimer dynamiquement la nécessité des maladies ou pannes, et une modélisation de la pertinence des actions possibles via la logique possibiliste pour déterminer si le comportement de l'opérateur est conforme aux règles métier.

Parmi les problématiques citées dans la section 6, nous avions noté que Dempster-Shafer suppose que tous les éléments de Ω sont mutuellement exclusifs, ce qui n'est pas toujours le cas dans les applications que nous envisageons. Des modifications du cadre ont été proposées pour prendre en compte la conjonctions de plusieurs hypothèses, par exemple dans [4], Laurence Cholvy considère que leur conjonction est une hypothèse que l'on ajoute explicitement au cadre. Cela permet de mieux représenter la réalité des problématiques auxquelles les opérateurs sont confrontés. Selon le nombre d'hypothèses considérées, le nombre d'états peut exploser.

De plus, la théorie des fonctions de croyances suppose un cadre de discernement fermé, comme nous l'avons évoqué dans la section 2 et la section 6. Nous faisons l'hypothèse que toutes les maladies possibles sont décrites dans Ω puis rassemblées dans $\mathcal{H}$. En pratique, ce n'est pas réaliste et nous voudrions conserver la possibilité d'autres maladies non représentées. Pour cela, il est possible de s'appuyer sur le modèle de croyances transférables [21] qui définit un cadre similaire à la théorie des fonctions de croyances tout en adoptant une hypothèse de monde ouvert, sans normalisation, où la masse sur l'ensemble vide représente soit un conflit réel, soit la possibilité que la vérité ne se situe pas dans le cadre considéré. Ce modèle pourrait théoriquement permettre de mieux représenter l'état du monde, en particulier hors de scénarios contrôlés, mais c'est une piste qui reste à explorer.

Enfin, nous avons souligné la problématique de l'attribution des masses dans la théorie des fonctions de croyances. Il faudra être en mesure de calibrer les valeurs du modèle informatique pour chaque scénario (à l'aide d'une méthode de calcul de point fixe ou d'algorithmes évolutionnaires d'exploration de l'espace des paramètres comme CMA-ES [10] qui est souvent utilisé en simulation) afin d'optimiser les différentes masses pour obtenir des résultats correspondant à ceux observés dans nos scénarios réels avec des médecins humains dans de bonnes conditions, c'est-à-dire sans biais.

Malgré ces limites bien identifiées, notre travail ouvre des perspectives intéressantes en termes de modélisation de l'erreur humaine. Nous voudrions compléter notre modèle en y intégrant une représentation des désirs et intentions des opérateurs, qui permettraient d'expliquer des actions qui ne sont pas nécessairement des erreurs (par exemple lorsque le médecin donne priorité à la survie du patient plutôt qu'à son diagnostic) et de prendre en compte les conditions et effets des actions comme formalisé dans les logiques d'action de type BDI. Nous voudrions aussi modéliser différents niveaux d'urgence (au lieu de seuils fixes pour l'ensemble des maladies) pour qu'une piste relativement peu probable mais grave, donc représentant un risque important, puisse être explorée par l'opérateur. Il pourrait au contraire être intéressant de ne produire certaines alertes qu'en situation d'urgence et ainsi de ne pas saturer l'opérateur, ce qui ris-

querait de le désensibiliser et de diminuer sa réceptivité. De plus, nous avons pu remarquer que certaines alertes produites dans nos expérimentations sont un peu trop rigides ; une action épistémique peu coûteuse pour tester une maladie ou une panne, même dont la masse est faible, n'est pas nécessairement une erreur.
Enfin, dans le cadre du projet IDEFIX, ce modèle sera utilisé sur différents scénarios, médicaux et aéronautiques, auprès de professionnels. Quatre scénarios en médecine ont été implémentés et quatre autres en aviation sont en cours d'implémentation. Nous modéliserons les activités des 30 sujets dans chaque domaine (aviation et médecine) à l'aide de notre modèle pour étudier l'apparition d'alertes et déterminer comment calibrer notre modèle en vue d'expérimentations futures : les verbalisations des médecins lors des scénarios seront alors des outils précieux pour évaluer la pertinence des prédictions de notre modèle.
En appliquant notre approche à ces différents scénarios, nous pourrons vérifier si le cadre de modélisation que nous avons développé est suffisamment riche pour donner des résultats cohérents à la fois dans des contextes différents et dans plusieurs domaines métiers.

À terme, nous espérons construire un modèle qui permette non seulement d'alerter l'opérateur sur la possibilité d'un biais de fixation, mais aussi de produire des explications sur ces alertes, ce qui peut les rendre plus acceptables par l'humain. Notre objectif, dans une démarche d'interaction humain-IA, est de proposer un modèle d'IA explicable pour aider les médecins et les pilotes à mieux éviter ces erreurs qui peuvent avoir des conséquences dramatiques.

Références

[1] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook : Theory, Implementation and Applications*. Cambridge University Press, 2 edition, 2007.

[2] A Bertaux, B Alameda, J Tataw, and A Kenfak. Effet tunnel en contexte d'épidémie. *Revue Médicale Suisse*, 16(718) :2392–2396, 2020.

[3] Alex Chaparro, Joseph Keebler, Elizabeth Lazzara, and Anastasia Diamond. Checklists : A review of their origins, benefits, and current uses as a cognitive aid in medicine. *Ergonomics in Design : The Quarterly of Human Factors Applications*, 27, 01 2019.

[4] Laurence Cholvy. Non-exclusive hypotheses in dempster–shafer theory. *International Journal of Approximate Reasoning*, 53(4) :493–501, 2012.

[5] Alessandro Cimatti and Marco Schaerf. Abductive reasoning in description logics. *Journal of Automated Reasoning*, 22(1) :1–39, 1999.

[6] Asaf Degani and Earl L. Wiener. Cockpit checklists : Concepts, design, and use. *Human Factors*, 35(2) :345–359, 1993.

[7] Arthur P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38(2) :325–339, 1967.

[8] Didier Dubois and Henri Prade. *Possibility Theory : An Approach to Computerized Processing of Uncertainty*. Plenum Press, 1988.

[9] Evie Fioratou, Rhona Flin, and Ronnie Glavin. No simple fix for fixation errors : cognitive processes and their clinical applications. *Anaesthesia*, 65(1) :61–69, 2010.

[10] Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evol. Comput.*, 9(2) :159–195, June 2001.

[11] David Heckerman, Eric Horvitz, and Bharat Nathwani. The pathfinder system. *Proceedings / the ... Annual Symposium on Computer Application [sic] in Medical Care. Symposium on Computer Applications in Medical Care*, 11 1989.

[12] Daniel Kahneman and Amos Tversky. Judgment under uncertainty : Heuristics and biases. *Science*, 185(4157) :1124–1131, 1974.

[13] David Lyell and Enrico Coiera. Automation bias and verification complexity : A systematic review. *Journal of the American Medical Informatics Association*, 24 :ocw105, 08 2016.

[14] John McCarthy. Circumscription – a form of non-monotonic reasoning. *Artificial Intelligence*, 13(1–2) :27–39, 1980.

[15] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[16] David Poole. Probabilistic horn abduction and bayesian networks. *Artificial Intelligence*, 64(1) :81–129, 1993.

[17] Daniel J. Power. *Decision Support Systems : A Historical Overview*, pages 121–140. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[18] Eric Raufaste, Rui da Silva Neves, and Claudette Mariné. Testing the descriptive validity of possibility theory in human judgments of uncertainty. *Artificial Intelligence*, 148(1) :197–218, 2003. Fuzzy Set and Possibility Theory-Based Methods in Artificial Intelligence.

[19] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[20] Edward Shortliffe. Computer-based medical consultations : Mycin. *Artificial Intelligence - AI*, 388, 10 1976.

[21] Philippe Smets and Robert Kennes. The transferable belief model. *Artificial Intelligence*, 66(2) :191–234, 1994.

[22] World Health Organization. *WHO Surgical Safety Checklist and Implementation Manual*. World Health Organization, Geneva, 2009. First Edition.

[23] Lotfi A. Zadeh. A simple view of the dempster-shafer theory of evidence and its implication for the rule of combination. *AI Magazine*, 7(2) :85, Jun. 1986.