

*BULLETIN*

*DE*

*L'AFIA*



**JANVIER 2013**

**N° 79**

*Association Française pour l'Intelligence Artificielle*

## Présentation du bulletin

Le Bulletin de l'Association Française pour l'Intelligence Artificielle vise à fournir un cadre de discussions et d'échanges au sein de la communauté universitaire et industrielle. Ainsi, toutes les contributions, pour peu qu'elles aient un intérêt général pour l'ensemble des lecteurs, sont les bienvenues. En particulier, les annonces, les comptes rendus de conférences, les notes de lecture et les articles de débat sont très recherchés. Le Bulletin de l'AFIA publie également des dossiers plus substantiels sur différents thèmes liés à l'IA. Le comité de rédaction se réserve le droit de ne pas publier des contributions qu'il jugerait contraire à l'esprit du bulletin ou à sa politique éditoriale. En outre, les articles signés, de même que les contributions aux débats, reflètent le point de vue de leurs auteurs et n'engagent qu'eux-mêmes.

### Pour contacter l'AFIA

#### Président

Yves DEMAZEAU

L.I.G./C.N.R.S., Maison Jean Kuntzmann, 110, avenue de la Chimie, B.P. 53, 38041 Grenoble cedex 9

Tel : +33 (0)4 76 51 46 43

Fax : +33 (0)4 76 51 49 85

[Yves.Demazeau@imag.fr](mailto:Yves.Demazeau@imag.fr)

<http://membres-lig.imag.fr/demazeau>

#### Serveur WEB

<http://www.afia.asso.fr>

#### Adhésions, liens avec les adhérents

Thomas GUYET

Laboratoire Informatique d'Agrocampus-Ouest

65, rue de Saint-Brieuc

35042 Rennes cedex

Mél. : [tresorier@afia.asso.fr](mailto:tresorier@afia.asso.fr)

#### Membres d'honneur

Marie-Odile Cordier (1999), Jean-Paul Haton (1999), Jacques Pitrat (1999), Jean-Marc David (2000), Daniel Kayser (2000), Claude Vogel (2000), Henri Farreny (2001), Alain Colmerauer (2002), Jean-Louis Laurière (2002), Gérard Sabah (2003), Jean-Claude Latombe (2004), Yves Kodratoff (2004), Malik Ghallab (2005), Marie-Christine Rousset (2005), Christian Bessière (2006), Luis Farinas del Cerro (2006), Pierre Marquis (2009), Jérôme Lang (2009), Michèle Sebag (2010), Jean-Gabriel Ganascia (2010), Serge Dupuy (2011).

#### Personnes morales adhérentes à l'AFIA

ENSMSE, Université Paris Dauphine, LORIA, LIRIS, LIMSI, IRIT/SMAC, EDF/STEP, LIPADE, IFFSTAR, LIRMM, TAO, LIFL, GREYC, LIG, ONERA, IRSTEA-TETIS, INRA, LITIS

#### Conseil d'Administration de l'AFIA

Yves DEMAZEAU, président

Jean-Denis MULLER, vice-président

Olivier BOISSIER, secrétaire

Catherine TESSIER, secrétaire adjoint

Thomas GUYET, trésorier

Patrick REIGNIER, webmestre

Amélie CORDIER, webmestre adjoint

Membres :

Patrick ALBERT, Christine BOURJOT, Catherien FARON-ZUCKER, Serge GARLATTI, Sébastien KONIECZNY, Vincent LEMAIRE, Nicolas MAUDET, Philippe MORIGNOT, Jean ROHMER, Laurent VERCOU-TER, Pierre ZWEIGENBAUM.

## Comité de Rédaction

**Laurence Capus**

**Rubrique « I.A. au Québec »**

Dept. d'Informatique et de Génie Logiciel

Université Laval, Québec, Canada

[laurence.capus@ift.ulaval.ca](mailto:laurence.capus@ift.ulaval.ca)

**Nicolas Maudet**

**Rubrique « I.A. sur les blogs »**

LIP6, Université Pierre et Marie Curie

4, place Jussieu, 75005 Paris

[maudetn@lip6.fr](mailto:maudetn@lip6.fr)

**Philippe Morignot**

**Rédacteur en chef**

IMARA, INRIA Rocquencourt

Domaine de Voluceau, B.P.105, 78150 Le Chesnay

[pmorignot@yahoo.fr](mailto:pmorignot@yahoo.fr)

**Patrick Reignier**

**Rubrique « Résumés de thèse et HDR »**

PRIMA, INRIA Rhône-Alpes

655, avenue de l'Europe, 38334 Saint-Ismier cedex

[Patrick.Reignier@inrialpes.fr](mailto:Patrick.Reignier@inrialpes.fr)

**Laurent Vercoüter**

**Rédacteur en chef adjoint**

LITIS, INSA de Rouen

avenue de l'université, BP8 76801 St-Étienne-du-Rouvray

[laurent.vercoüter@insa-rouen.fr](mailto:laurent.vercoüter@insa-rouen.fr)



## Dossier Éthique et I.A.

Comment concilier autonomie d'un système logiciel ou robotique, et enjeux sociétaux tels que la protection des données individuelles ? Le concept sous-jacent à cet interrogation est celui de l'éthique : le dossier de ce trimestre est consacré à l'I.A. et éthique, préparé par Olivier Boissier (Ecole des Mines de Saint-Etienne) et Catherine Tessier (ONERA Toulouse). Un panorama des équipes travaillant sur ces thèmes sont recensés par les deux coordonnateurs, qu'ils en soient ici remerciés.

Ce dossier inclut deux comptes-rendus de journées relatives à ce thème, lors de RFIA'12 et ECAI'12.

Enfin, nous publions le résumé de l'article primé aux JFSMA'12, rédigé par Benoit Romito et François Bourdon (Université de Caen).

Rappelons également le lien sur le site de notre association, où foisonnent annonces, cours et ressources en général :

<http://www.afia.asso.fr>

Bonne lecture.

**Philippe Morignot & Laurent Vercouter**  
Rédacteurs en chef

## Dossier Éthique et I.A.

Avec la présence croissante des technologies numériques au sein des différentes facettes de notre société, l'éthique des STIC est devenue une question cruciale. On en parle dans la presse ; elle devient alibi pour attiser les peurs du grand public ; des comités, des commissions de réflexion, des groupes de travail sont créés.

Cette question va au-delà des réflexions relatives à la protection de la vie privée et liées à la généralisation de la collecte, de l'exploitation et de la conservation de masses de données individuelles. Elle aborde le nouvel enjeu de l'autonomie des systèmes, enjeu de plus en plus central dans la conception de systèmes actuels. En effet, la décision artificielle se trouve au cœur d'applications telles que le commerce électronique, les jeux interactifs, l'intelligence ambiante, l'assistance aux personnes, la robotique sociale ou collective, les systèmes de véhicules sans pilote.

De nombreuses questions à la fois d'ordre technologique mais aussi relatives à l'usage et aux impacts sociétaux se posent. Comment concilier autonomie et comportement éthique dans des agents artificiels et dans des systèmes hybrides mêlant agents logiciels ou robotiques et humains ? Quelle(s) éthique(s) considérer ? Comment concevoir et mettre en œuvre des règles qui relèvent de la philosophie, du bon sens, des valeurs, qui dépendent du contexte, qui peuvent être contradictoires ? Et de manière plus générale, comment spécifier une IA éthique et quelles recommandations faire aux chercheurs s'attaquant à ces problématiques ?

Si de telles questions doivent être posées et traitées en amont de chaque projet de recherche en Intelligence Artificielle, encore assez peu d'équipes en France y travaillent spécifiquement. Dans ce numéro, nous proposons les contributions de certaines d'entre elles. Nous aurions pu l'intituler « IA Ethique, Éthique de l'IA » tant il aborde ces questions selon le point de vue de l'usage des technologies ou selon le point de vue des technologies. Nous avons également ajouté les comptes rendus des ateliers « Agents Autonomes et éthique », organisé dans le cadre de RFIA 2012, et « RDA2 Workshop on Rights and Duties of Autonomous Agent », organisé dans le cadre de ECAI 2012. Les contributions présentées dans ces deux ateliers constituent autant d'éléments pour poursuivre

la réflexion sur ces questions essentielles.

Olivier Boissier et Catherine Tessier

## Éthique & IA : de l'éthique appliquée à la méta-éthique

Avec le développement de l'intelligence artificielle (IA) et, plus généralement, des technologies de l'information et de la communication (TIC), le monde change. Le travail se transforme. Les notions de souveraineté et d'autorité évoluent. Les règles de vie en société se modifient. Dans quelle mesure les fondements sur lesquels reposent les commandements moraux en sont-ils affectés ? Telle est la question que nous traiterons ici en envisageant tour à tour les trois branches traditionnelles de l'éthique, à savoir l'éthique appliquée, l'éthique normative et la méta-éthique

- CONTACT : Jean-Gabriel Ganascia, [jean-gabriel.ganascia@lip6.fr](mailto:jean-gabriel.ganascia@lip6.fr),
- ADRESSE : équipe ACASA - LIP6 - Université Pierre et Marie Curie, B.C. 169, 4 place Jussieu, 75005 Paris
- WEB : <http://ganascia.name>
- TEL : +33 1 44 27 37 27

### L'étable

Le mot « éthique » provient du grec *ethos* qui désigne les mœurs, c'est-à-dire les habitudes de comportement relativement au bien et au mal. En cela, il se rapproche du mot « morale » (qui vient de *mores*, mœurs en latin), quand bien même il existerait de subtiles distinctions entre éthique et morale. Celles-ci tiennent sans doute aux différences entre la Grèce et Rome : d'un côté, une culture philosophique éprise de ratiocination ; de l'autre, une civilisation juridique, soucieuse de règles et de lois. Ainsi, la morale énonce les règles de conduite, tandis que l'éthique cherche à établir le fondement de ces règles.

Plus déconcertante sans doute est l'étymologie d'*ethos* en grec : avant de désigner les coutumes, ce mot signifiait l'*étable*, à savoir l'endroit où les bestiaux rentrent tous les soirs, c'est-à-dire, le lieu de leur identité. L'éthique renvoie donc à la durée. En effet, seul celui qui perdure peut répondre de ses actes. Quant aux coutumes auxquelles l'*ethos* fait aussi référence, elles impliquent toutes une transmission dans le temps.

Dans un monde aussi changeant que le nôtre, l'éthique imposerait donc une continuité : elle énoncerait ce qui demeure en dépit des changements consécutifs au déploiement des techniques en général, et des TIC en particulier, en dépit des transformations du travail, en dépit de l'évolution des formes politiques due à la

« dé-territorialisation », c'est-à-dire à la dissociation de l'état et du territoire, etc. Il y aurait donc une tension entre l'IA, vecteur d'évolution, et l'éthique, en quête de pérennité. Or cette tension nous apparaît féconde non seulement pour l'IA, afin que les techniques produites ne contrarient pas nos valeurs et qu'elles répondent aux aspirations humaines, mais aussi pour l'éthique, susceptible de se renouveler grâce à l'IA.

Aujourd'hui, personne ne doute que l'éthique apporte un « supplément d'âme » aux techniciens, tout en suscitant de nouveaux développements comme, par exemple, les architectures de logiciels qui préservent la vie privée, aussi, nous ne nous étendrons pas sur ce point. En revanche, que l'IA contribue à l'éthique surprend plus. C'est donc sur ce dernier point que portera cette contribution.

### Divisions de l'éthique

Pour prendre la mesure des apports de l'IA à l'éthique, soulignons d'abord que les philosophes distinguent aujourd'hui trois secteurs de l'éthique. Le premier, indubitablement le plus concret, prend en compte l'objet particulier sur lequel portent les conduites, par exemple les patients pour les médecins, le climat pour les climatologues, l'information pour les informaticiens, etc. Cela correspond à ce que l'on appelle l'« éthique appliquée ». Le second porte sur les fondements des règles de conduite ; on le range sous le vocable de « éthique normative ». Enfin, le dernier volet désigné parfois comme la « méta-éthique » n'énonce pas de principes à proprement parler, mais il précise la signification des énoncés moraux et, surtout, leur statut aux plans logique, psychologique, métaphysique, juridique etc. Or, comme nous allons le voir dans la suite, ces trois secteurs apparaissent influencés par les développements récents de l'IA.

### Un nouveau secteur de l'éthique appliquée

En premier lieu, un nouveau secteur de l'éthique appliquée porte désormais sur les conséquences du développement des TIC et en particulier de l'IA. Il apparaît évident aujourd'hui que toutes ces techniques pénètrent profondément dans notre vie quotidienne et qu'elles modifient le travail, les relations humaines, les communications, l'économie, etc. Bref, les mœurs, au sens de coutumes et d'usages des hommes en société se transforment. Il s'ensuit que les conditions de vie changent, que de nouvelles vulnérabilités apparaissent et, en conséquence, que les notions de bien et de mal demandent à être repensées

et, surtout, que des questionnements inédits se font jour. à titre d'illustration, prenons trois exemples.

#### *Transparence et protection de la vie privée*

Le premier porte sur l'accès aux données individuelles. Indubitablement, leur exploitation bénéficiant grandement de l'apport des techniques d'IA. Parallèlement, le déploiement massif des réseaux et des ordinateurs conduit à l'accumulation de plus en plus conséquente d'informations de toutes sortes, par exemple, des traces de parcours sur le web, des courriels échangés ou des photographies déposées sur les réseaux sociaux. Dès lors, l'individu risque de se voir traqué du fait d'informations laissées à son insu, de nombreuses années auparavant, sur un site ou un blog. Introduite au XIX<sup>e</sup> siècle, au moment où la presse populaire commençait à se diffuser largement, l'exigence de protection de la vie privée semble donc de plus en plus actuelle.

Parallèlement, l'exigence de transparence devient de plus en plus prégnante dans les démocraties modernes ; on demande aux hommes publics, dès qu'ils exercent un pouvoir quelconque, par exemple, celui de gouverner, d'administrer, d'enseigner, de soigner, de juger, etc., de livrer toute information les concernant, afin de prouver leur neutralité et leur moralité. Et les technologies actuelles autorisent un enregistrement permanent de toutes les actions individuelles. Il s'ensuit que le désir de protéger sa vie privée entre en conflit avec la revendication légitime de transparence qui conduirait à exploiter toutes les sources disponibles d'information sur les individus [17]... Comment arbitrer entre ces deux besoins ? Ce sera certainement la fonction d'un champ spécifique de l'éthique appliquée que de nous aider à y répondre.

#### *Conservation des données*

Le second exemple porte sur la conservation des données. Aujourd'hui, beaucoup déplorent la fraude scientifique [14]. Introduit aux États-Unis en 2000 par l'OSTP (*Office of Science and Technology Policy*), le sigle « FFP » caractérise les trois grandes catégories de fraudes : *Fabrication, Falsification, Plagiat*. Pour s'assurer que les données ne sont pas fabriquées et/ou falsifiées, le scientifique doit, en toute rigueur, les conserver après la publication des résultats, afin d'autoriser des vérifications ultérieures. Or, si ces données concernent des individus, par exemple s'il s'agit de données médicales ou de données d'usage comme les enregistrements de nos parcours sur Internet, le scientifique doit absolument s'engager à les détruire après son expérimentation... Là encore,

nous constatons des tensions inédites entre des exigences morales conflictuelles. Cela indique, à l'évidence, la nécessité d'un nouveau champ de l'éthique appliquée pour tenter de les résoudre.

#### *Responsabilité*

Songez maintenant au devenir de la notion de responsabilité. Pendant longtemps, la morale exigeait de l'homme qu'il anticipe les conséquences de ses actions. Or, aujourd'hui, l'outil devient si complexe que tant le concepteur que le fabricant, le propriétaire ou l'utilisateur, ne saurait en garantir le comportement en toutes circonstances. En d'autres termes, l'homme ne saurait répondre, à tout moment, des conséquences de ses actes, dès lors qu'ils impliquent l'échange avec des robots autonomes. La notion de responsabilité évolue donc considérablement lorsque les hommes se trouvent confrontés à des dispositifs qualifiés d'« intelligents », parce que l'on pourrait imaginer qu'une entité intelligente les anime. Là encore, l'éthique appliquée à l'IA trouve de nouvelles problématiques à aborder.

#### **éthique normative - apports de la formalisation**

Cependant, comme nous l'avons suggéré plus haut, l'apport de l'IA à l'éthique ne se limite pas à la constitution d'un nouveau secteur de l'éthique appliquée, loin de là. En effet, l'éthique normative en bénéficie déjà et pourrait en bénéficier plus encore. À cet égard, rappelons d'abord que l'éthique normative porte sur le fondement des règles de conduite. Par exemple, pourquoi ne doit-on ni voler, ni mentir ? Est-ce parce qu'une loi nous l'interdit ? Parce que mon intérêt bien compris s'y oppose ? Parce que cela m'avilirait ? Ou encore, parce que des personnes pourraient en pâtir ? Bien évidemment, les commandements éthiques varient selon les justifications adoptées.

Dans la mesure où l'éthique normative décrit des systèmes cohérents de justifications, elle doit pouvoir faire l'objet de formalisations. Depuis longtemps, les logiciens font appel à des logiques modales particulières, dites *déontiques* [45], parce qu'elles formalisent les devoirs et les droits. Néanmoins, ces formalismes se heurtent à des contradictions lorsque les commandements se contredisent [30, 20]. Or, l'éthique permet essentiellement de surmonter les conflits ; s'il n'y en avait pas et s'il suffisait d'appliquer des règles de morale, elle ne servirait à rien. La notion de non-monotonie a été introduite en intelligence artificielle pour surmonter les contradictions. De ce fait, elle semble parfaitement adaptée à la modélisa-

tion des conflits éthiques [21]. Plusieurs travaux ont été poursuivis dans cette perspective depuis plusieurs années [34, 22, 16]. L'intérêt de ces travaux tient non seulement à la possibilité de résoudre des dilemmes éthiques, mais aussi, et surtout, à formaliser différents systèmes normatifs puis à les comparer, ce qui faisait justement l'objet de l'éthique normative. Bref, indubitablement, celle-ci est en passe de se renouveler avec l'IA.

### Vers un renouveau de la méta-éthique

Dernier point, l'IA affecte aussi la méta-éthique. Sans entrer dans le détail, rappelons que cette branche de l'éthique vise à préciser le statut des énoncés moraux, sur tous les plans, en particulier aux plans logique, métaphysique, psychologique, etc. Or, avec les récents développements de l'IA, la signification d'énoncés éthiques évolue grandement. À titre d'illustration, on parle de plus en plus de robots ou d'agents artificiels « éthiques » [5]. Certains ont même imaginé un test de Turing « éthique » [46] qui viserait à identifier non le sexe, comme dans le test de Turing original [41], mais la moralité d'un agent artificiel. Or l'attribution de qualificatifs moraux à des machines ne va pas sans remettre en question la signification même des énoncés moraux, qui, jusque là, ne s'appliquaient qu'à des êtres doués d'une conscience. L'examen de ces évolutions relève donc indubitablement de la méta-éthique. D'ailleurs, certains philosophes [36] vont même jusqu'à évoquer un « droit des machines » analogue au « droit des animaux » revendiqué il y a quelques années par d'autres philosophes [35]. Certes, on ne songe pas ici à réduire l'éthique à un jeu de règles formelles. De même qu'avec le *Knowledge Level* [33], Alan Newell attribuait une connaissance à des machines, sans la réduire à son expression dans un langage quel qu'il soit, on pourrait imaginer approcher le caractère éthique d'un agent artificiel, sans le réduire aux règles de conduite qu'il utilise pour réguler son comportement. Sa moralité se révélerait alors non pas dans ses propres actions, mais dans les dialogues et les échanges auxquels il nous convie et qui nous invitent à réfléchir aux conséquences de nos actes et à leur dimension éthique, voire à modifier nos comportements en conséquence. Nous serions alors amenés à étendre la portée de l'éthique à des formes d'existence non conscientes. Cette extension relève de la méta-éthique au sens où nous l'avons définie, car elle attribue un statut à des énoncés éthiques portant sur des agents artificiels, alors qu'auparavant, ceux-là en étaient dénués.

### Retour à l'étable...

Sans doute, avec le développement de l'IA et, plus généralement, des technologies de l'information, le monde change; la condition humaine évolue; les mœurs aussi. Nos façons d'aborder les problèmes diffèrent. Nous n'employons plus les mêmes mots pour nous exprimer. Et, démultipliée par la technique, la portée de nos paroles et de nos actes s'étend considérablement, ce qui accroît nos responsabilités, tandis que nous maîtrisons de moins en moins les conséquences de nos décisions. Mais les questions demeurent. Et, avec elles, l'exigence de rumination subsiste aussi forte, si ce n'est plus qu'auparavant. Or, l'éthique nous offre une étable où nous sommes en mesure d'accomplir ce devoir de rumination...

## Équipe IRA2

Les SMA's seraient-ils les futurs garants de l'éthique en IA ?

- CONTACT : Nadia Abchiche-Mimouni - [nadia.abchiche@ibisc.univ-evry.fr](mailto:nadia.abchiche@ibisc.univ-evry.fr)
- WEB : <http://www.ibisc.univ-evry.fr/ira2>
- TEL : +(33)1 64 85 34 74

### Membres permanents de l'équipe

- CHELLALI Amine, Maître de Conférences
- COLLE étienne, Professeur
- HOPPENOT Philippe, Professeur
- MALLEM Malik, Professeur
- OTMANE Samir, Professeur
- ABABSA Fakhreddine, Maître de Conférences, HDR
- ABCHICHE-MIMOUNI Nadia, Maître de Conférences
- BENCHIKH Laredj, Maître de Conférences
- BOUYER Guillaume, Maître de Conférences
- DIDIER Jean-Yves, Maître de Conférences
- HADJ-ABDELKADER Hicham, Maître de Conférences
- MONTAGNE Christophe, Maître de Conférences
- ROUSSEL David, Maître de Conférences

### Thème général de l'équipe

L'originalité de l'équipe IRA2 est de considérer les enjeux sociétaux comme source d'inspiration pour les problématiques scientifiques qu'elle adresse et les méthodes qu'elle explore. Ainsi, ses recherches portent sur

la conception, la réalisation et l'évaluation de systèmes d'assistance à la personne dans son environnement intérieur ou extérieur. La diversité, la complexité et l'imprévisibilité de l'environnement nécessitent l'utilisation et la maîtrise de systèmes complexes artificiels (capteurs et robots hétérogènes). Les problématiques scientifiques se situent au niveau de la perception, de l'interaction et de l'aide à la décision à la personne. Il s'agit de proposer :

- des assistances multisensorielles et collaboratives à l'interaction et à la prise de décision à la personne, en situation d'immersion ou de semi immersion dans un monde virtuel,
- des assistances à la personne en interaction directe avec l'environnement réel augmenté,
- des fonctionnalités à un robot assistant coopérant avec l'opérateur et évoluant dans un environnement domotisé.

### Description des travaux

Le travail présenté ici est le fruit de plusieurs recherches dont les résultats sont publiés dans [1], [2] et [10]. Nous discuterons l'intégration de la dimension éthique qui doit dépasser le simple fait de considérer la gêne occasionnée, l'acceptabilité ou le respect de la vie privée des personnes. On décrira le contexte applicatif de nos travaux, on montrera les résultats obtenus avec une approche multi-agents à base de mécanismes de coalitions, et enfin, on mènera une réflexion sur une vision d'une éthique intégrative.

#### Contexte applicatif et sociétal

L'assistance à la personne est un des domaines de l'informatique dans lequel la dimension éthique s'invite d'elle-même. Pas un projet n'est financé sans un partenaire des sciences humaines et sociales. Les questions du respect de la vie privée et de la gêne occasionnée par les dispositifs installés à proximité des personnes et l'acceptabilité de ces dispositifs, font désormais partie intégrante de ces projets.

#### Problématique

Nous nous sommes intéressés à des scénarios où un robot et plusieurs objets communicants présents dans la résidence d'une personne dépendante participent à la levée de doute sur une alarme. La réussite de la mission dépend de la capacité du robot à se déplacer d'un point de départ vers une destination de façon fiable, et ce, quelles que soient les situations rencontrées. Les objets commu-

nicants de l'environnement peuvent jouer un rôle *facilitateur* en aidant le robot d'assistance à se localiser, à naviguer, ou à rechercher la personne. Inversement, le robot est vu comme un objet communicant mis à contribution par des services à la personne tels que la stimulation cognitive ou de télésurveillance.

De nombreux enjeux scientifiques peuvent être soulevés par ce type de scénarios. Les trois questions principales auxquelles nous avons cherché à apporter des éléments de réponse sont :

- La localisation du robot en environnement intérieur par coopération avec un environnement ambiant. Comment obtenir un résultat de localisation garanti, sous certaines hypothèses, en présence d'un réseau de capteurs hétérogènes dont le nombre et les caractéristiques sont variables? L'idée est que la connaissance fiable de sa localisation, même si celle-ci est dégradée (imprécise), permettra au robot d'adapter sa stratégie de déplacement en fonction de la situation.
- L'adaptation au contexte. Quel(s) mécanisme(s) d'adaptation pour gérer : i) la variété et la disponibilité des capteurs, ii) les contraintes liées au service rendu à la personne (qualité de service, degré d'urgence) ?
- La dimension éthique. Comment concevoir un système qui soit capable de respecter la vie privée des personnes tout en portant assistance à la personne (gestion d'un niveau d'intrusion) ?

Nous avons proposé une architecture informatique permettant la coopération entre le robot et l'environnement ambiant de façon adaptative. Cette architecture, basée sur un système multiagent (SMA) adaptatif et une ontologie, est décrite ci-après.

#### Architecture de coopération

La Figure 1 montre l'architecture globale du système dont les composantes sont : une base de connaissance orientée domaine d'application, un SMA qui prend en charge de manière adaptative le caractère dynamique et hétérogène de l'environnement ambiant et une passerelle en charge de l'uniformisation des informations issues de l'environnement ambiant pour la perception des agents. Chaque objet communicant de l'environnement est encapsulé par un agent. Les principes qui sous-tendent le mécanisme d'adaptation du SMA sont à base de mécanismes de coalitions, d'où son nom, COALAA. Les coalitions d'agents ont pour objectif de répondre à une requête d'un effet souhaité. La notion d'effet est vue comme un service de granularité variable qui doit être fourni par

le système à l'humain. Ce dernier peut être la personne en perte d'autonomie ou un de ses proches, un producteur de soin ou un administrateur distant. On peut citer comme effets souhaités la localisation du robot, l'éclairage d'une zone particulière d'une pièce de l'habitation, ou encore la recherche de la personne.

#### *Mécanismes d'adaptation*

étant donné la complexité d'une approche centralisée dont l'objectif serait de trouver une configuration des objets communicants de l'environnement qui réponde à un effet souhaité tout en prenant en compte l'aspect dynamique, notre approche à base de coalitions d'agents a fourni des résultats prometteurs. En effet, notre algorithme de formation de coalitions prend en compte l'aspect dynamique et évolutif de l'environnement et gère les critères et contraintes sur l'effet souhaité, de manière dynamique. Ses principales étapes sont décrites dans [2]. C'est l'étape de négociation qui gère la relaxation des critères qui permet d'assurer une solution (même dégradée) dans la majorité des situations.

#### *Prise en compte du niveau d'intrusion*

Le niveau d'intrusion modélise, sur une échelle de valeurs, la gêne occasionnée sur la personne par les objets communicants et le robot. Ce paramètre permet de moduler les actions des capteurs ou des actionneurs de l'environnement. Ainsi, le système fonctionne initialement avec un niveau d'intrusion minimal. Si au bout d'un temps donné, le SMA n'a pas réussi à former une coalition pour obtenir l'effet souhaité, le niveau d'intrusion est incrémenté. Cela augmente le degré de mouvement des objets communicants, permettant par exemple, à une caméra mobile de s'actionner pour changer d'orientation. Le processus de formation de coalitions est relancé avec de nouvelles données ainsi acquises. Enfin, le niveau d'intrusion maximal autorise le robot à se déplacer dans l'habitation du patient.

### **Résultats et discussion**

La validation de COALAA est publiée dans [2]. Une gestion à base de règles permettant une modification dynamique des critères a été implantée et est publiée dans [1]. Dans cette dernière version, l'utilisateur peut configurer les critères à prendre en compte, lors de la formation des coalitions et de la gestion de la priorité sur ces critères, par simple envoi de messages aux agents. Notons que les critères en question concernent aussi bien

les aspects techniques que des aspects éthiques (niveau d'intrusion et degré d'urgence). La prise en compte de critères éthiques au sein du système est en soi une avancée et est rendue naturelle par les mécanismes adaptatifs de COALAA. Cependant, l'éthique dépasse le simple fait de paramétrer des critères ; des questions fondamentales demeurent posées. Par exemple, l'intervention d'un robot d'assistance pose des questions sur l'impact de la compagnie d'un robot sur l'évolution des notions d'altérité [39] et de culture [38]. Nous soutenons que, pour être éthiques, les systèmes intelligents doivent être capables de raisonner explicitement sur l'éthique. Cela peut être rapproché de la notion de "collective awariness" développée dans [47].

#### *Une écologie de l'action*

Selon Edgar Morin ([31]), l'autonomie des entités disloque l'éthique car la spécialisation et le cloisonnement des connaissances tendent à morceler et diluer les responsabilités. Selon lui, l'éthique est pour les individus autonomes et responsables, l'expression de l'impératif de la reliance. Il définit une écologie de l'action comme une composante essentielle de l'éthique qui obéit à deux principes :

1. Importance du contexte : les effets de l'action dépendent non seulement de l'intention de l'acteur, mais aussi du contexte dans lequel elle se déroule.
2. Imprédictibilité à long terme : même si l'intention tente d'envisager les conséquences des actions, la difficulté à prévoir ces actions et l'incertitude dans la relation intention-action demeurent.

#### *Droits et devoirs des agents autonomes*

Dans les SMA, la question des normes, droits et devoirs des agents font l'objet de nombreux travaux dont la plupart ont pour objectif de garantir un comportement des agents en accord avec des règles prédéfinies. Citons en exemple les travaux présentés dans [48].

#### *Vers une vision intégrative de l'éthique*

La science moderne a exigé la séparation entre la connaissance et l'éthique, mais ces dernières décades, le développement de l'intelligence artificielle (IA) nous a poussés à une reconsidération. En effet, si l'IA à ses débuts était cloisonnée dans une machine à penser, elle est de plus en plus présente dans l'action et la décision, en interaction avec l'humain. Le développement d'agents

mobiles (logiciels ou robots) autonomes, proactifs, auto-adaptatifs a donné lieu à d'innombrables études psychologiques, anthropologiques et sociologiques montrant le fort impact de tels systèmes sur l'évolution de l'Homme. Cette évolution à son tour ayant un impact sur le développement des applications intelligentes, nous entrons dans le domaine de la pensée complexe [32].

La seule application de règles, aussi éthiques soient-elles, ne suffit pas à concevoir des systèmes réellement éthiques. Il faut non seulement expliciter l'intention de l'action, mais s'intéresser aussi aux effets et conséquences de l'action. En s'appropriant la notion d'écologie de l'action, l'idée est que les agents intègrent l'effet de leurs actions et l'incertitude quant à la prédiction de ces actions sur l'environnement, pour que le SMA puisse obéir aux deux principes énoncés en 1.4.1. étant donné la capacité intrinsèque des agents à percevoir et à agir sur leur environnement, l'implantation de ces deux principes est aisée. Il suffit pour cela que cet environnement soit ouvert dans le sens employé par Mario Tokotro [40]. De plus, la notion de reliance introduite par E. Morin, comme étant essentiel à l'éthique, peut être directement mise en correspondance avec le potentiel d'interaction inhérent aux SMA.

Les systèmes adaptatifs, autonomes et pro-actifs posent de réelles questions d'éthique, ils nous obligent à reconsidérer cette dimension. Paradoxalement, ils semblent être des supports privilégiés pour une prise en compte de l'éthique, en rejoignant l'étude d'Edgar Morin développée dans [31] et [32]. À ce titre, il nous paraît légitime de poser la question suivante : les SMA seraient-ils les futurs garants d'une éthique de l'IA ?

### Projets de l'équipe IRA2 en lien avec le travail décrit

Les projets CompanionAble (FP7 IP, 2008-2012) et QuoVADis (ANR TecSan, 2008-2011) s'intéressent au maintien à domicile des personnes dépendantes. Ils proposent, analysent et valident des systèmes et des techniques d'assistance selon deux axes : la sécurité de la personne et la stimulation cognitive.

## Équipe UsIA, CEA LIST

Ethique des systèmes autonomes

- CONTACT : Michaël Aupetit - [michael.aupetit@cea.fr](mailto:michael.aupetit@cea.fr)

- ADRESSE : LADIS, CEA LIST, Centre CEA de Saclay, Bat. 565, PC 192, 91191 Gif-sur-Yvette
- WEB : <http://www.michael.aupetit.free.fr>
- TEL : 01 69 08 88 61

### Membres de l'équipe

- Michaël Aupetit, chercheur, HDR Univ. Paris Sud 11, visualisation analytique et apprentissage automatique
- Laurence Boudet, chercheur, systèmes d'inférence floue
- Laurence Cornez, chercheur, systèmes d'inférence floue
- Javier Gil-Quijano, chercheur, systèmes multi-agents
- Jean-Philippe Poli, chercheur, apprentissage automatique et optimisation
- Nicolas Heulot, doctorant, visualisation analytique
- Maxime Maillot, doctorant, apprentissage automatique

### Thème général de l'équipe

Visualisation analytique, Systèmes multi-agents, Systèmes d'inférence floue et Apprentissage automatique, pour la modélisation de systèmes complexes orientée utilisateur

### Description des travaux

Nous travaillons sur la modélisation de systèmes complexes pour l'aide à la décision. Nous concevons des modèles permettant la fusion des connaissances issues des données avec celles des utilisateurs experts ou non, afin d'en faciliter l'acquisition et la restitution. Pour cela, nous combinons des techniques de visualisation analytique [3, 4, 25], des systèmes d'inférence floue [24] et d'apprentissage automatique [29, 15] et des systèmes multi-agents.

### *Accompagner l'utilisateur*

Notre confrontation avec l'utilisateur (expert ou non) de nos systèmes d'aide à la décision nous conduit à nous interroger sur les implications applicatives des questions d'éthique. Les systèmes d'aide à la décision servent d'intermédiaire entre les experts concepteurs du système, les experts fournisseurs de connaissances et les experts ou non experts utilisateurs du système. L'éthique est censée

indiquer comment doivent se comporter et agir les humains les uns envers les autres. Lorsque ces actes ne sont plus directs mais passent par l'intermédiaire d'un système doté d'Intelligence Artificielle, se pose la question de la responsabilité des différents acteurs qui ont conçu ce système en choisissant des modèles et les algorithmes d'inférence et d'optimisation de leurs paramètres, et lui ont transmis des connaissances d'expert directement, par exemple sous forme de règles « si ... alors ... » ou d'ontologies, ou indirectement sous forme du choix des capteurs, variables et signaux supposés contenir la connaissance requise et finalement acquise empiriquement et intégrée par le système au cours du temps. De même, l'utilisateur de ces systèmes d'aide à la décision doit assumer sa propre charge de responsabilité et, face à un système qu'il ne maîtrise pas, il suit généralement deux lignes de conduite distinctes : la confiance aveugle en la décision suggérée par le système, transférant parfois à ses dépens et sans en avoir conscience sa charge de responsabilité sur les concepteurs et experts l'ayant mis au point, ou la défiance systématique conduisant à la non utilisation du système.

Nous cherchons à développer des systèmes d'aide à la décision qui par un encodage graphique [4] ou linguistique [24] rendent interprétables les connaissances sur lesquelles ils basent leurs décisions afin de permettre l'appréhension et l'apprentissage du fonctionnement du système par l'utilisateur et d'éviter sa confiance aveugle ou sa défiance systématique.

#### *éthique des systèmes autonomes*

Ces questions de responsabilité deviennent de plus en plus prégnantes avec l'augmentation du degré d'autonomie des systèmes d'IA, i.e. leur capacité à s'éloigner éventuellement de manière imprévisible du cadre de fonctionnement pour lequel ils ont été conçus. On peut penser que seuls les robots et systèmes dotés d'actionneurs « physiques » représentent un danger potentiel pour les utilisateurs (pensons aux robots industriels des chaînes de production automobile ou aux systèmes de pilotage automatique des avions), mais le trading algorithmique a montré récemment la capacité d'un système complètement virtuel d'achat et de vente automatisé de titres boursiers d'agir *in fine* de manière tangible et nuisible sur le monde réel. L'arrivée massive de robots et de systèmes d'IA domotique dans les foyers ou le domaine public au contact de personnes non expertes et non destinées à l'être, pose de manière plus aiguë ces questions d'éthique et de responsabilité, et les moyens d'y répondre sont encore balbutiants. Un fabricant d'aspirateurs automatiques peut contractuellement se dégager

de toute responsabilité en cas d'accident engendré par ses robots, mais il ne vendra ses produits que si ses clients sont convaincus qu'ils peuvent maîtriser le risque lié à leur utilisation, ce dont ils sont responsables. Un robot de la taille, du poids et de la force d'un homme adulte représente un danger potentiel bien plus important que celui de la taille d'un jouet. Ce danger est d'autant plus difficile à prévoir et à maîtriser que le comportement de ce robot dépend davantage de l'expérience acquise dans son environnement spécifique que de connaissances communes préprogrammées. Le développement technique de telles machines rendu possible par les avancées scientifiques, nous oblige à aborder et résoudre les questions d'éthique que ces machines posent si l'on souhaite leur déploiement et leur développement commercial.

Nous pensons qu'un moyen de parvenir à maîtriser ces machines de plus en plus autonomes voire peut-être un jour conscientes d'elles-mêmes, est de rendre interprétables les connaissances et les processus décisionnels qu'elles utilisent afin de pouvoir détecter, diagnostiquer et corriger leurs défauts tout au long de leur fonctionnement. Nous devons concevoir des techniques d'analyse et de diagnostic de ces systèmes d'IA qui relèvent de la psychologie et de la psychiatrie chez l'humain. Nous avons cependant un avantage par rapport à nos collègues psychologues ou psychiatres... nous sommes concepteurs de ces systèmes !

#### Projets

- Eritr@c : projet européen. Développement d'un système de contrôle de conteneurs maritimes par analyse neutronique, et système d'aide à la décision pour les douaniers [4]. [http://www.euritrac.org/Francais/French\\_version.htm](http://www.euritrac.org/Francais/French_version.htm)
- Descartes : projet FUI (2008-2012). Développement d'un prototype d'un système de commandement et contrôle pour la gestion de crise intégrant des fonctions d'analyse, de prédiction, d'aide à la décision et de formation. <http://www.cea.fr/technologies/projet-descartes>
- Mobisic : projet FUI (2008-2011). Développement d'un système de sécurisation mobile pour la gestion de crise.

## ONERA-DCSD-CD

éthique et partage de l'autorité pour les engins à autonomie variable

- CONTACT : Catherine Tessier - [catherine.tessier@onera.fr](mailto:catherine.tessier@onera.fr)
- ADRESSE : ONERA Centre de Toulouse, 2 avenue Édouard-Belin, 31055 Toulouse cedex 04
- TEL : 05 62 25 29 14

## Membres de l'équipe travaillant sur le thème

- Catherine Tessier, maître de recherches
- Florian Gros, doctorant

## Thème général de l'équipe

L'unité de recherche Conduite et Décision (CD) du département Commande des Systèmes et Dynamique du vol (DCSD) de l'ONERA mène des études et recherches visant à développer les capacités d'autonomie comportementale et décisionnelle sûres de fonctionnement pour des engins habités ou non (aéronefs, engins spatiaux, engins sous-marins, robots), évoluant seuls ou en équipe, et plus ou moins supervisés par un opérateur humain.

## Description des travaux

Les recherches se placent dans le contexte du partage de l'autorité entre un agent robotique et un opérateur humain. L'accent est mis sur le contrôle de robots armés dont plusieurs fonctions (tir, détection de cibles, navigation, etc.) peuvent être autonomes. L'objectif est de pouvoir assurer la conformité des actions du système homme-robot aux directives du Droit International Humanitaire et des règles d'engagement spécifiques à la mission dans le cadre de situations moralement difficiles à résoudre. On cherche également à limiter les conflits d'autorité entre les agents, de façon à ce que le système ne réalise pas d'action imprévue ou inacceptable.

Pour cela nous étudions les aspects de confiance, d'autonomie et d'autorité, ainsi que les problèmes que ces notions posent au regard du robot lui-même et de l'interaction entre l'humain et l'agent robotique. Nous étudions de surcroît les implications du déploiement des systèmes armés autonomes sur le champ de bataille, ainsi que les moyens d'implémenter les différentes directives du Droit International Humanitaire au sein de ces systèmes [19, 18].

Des protocoles expérimentaux sont en cours de définition, avec l'objectif d'étudier comment un opérateur réagit à un dilemme lors d'une mission, quels critères sont pertinents pour le raisonnement moral dans de telles situations et quelles sont les métriques à évaluer lorsque le système homme-robot est mis face à un dilemme. Au

cours des expériences, on fera varier en particulier la manière dont les robots sous supervision de l'opérateur sont équipés de « règles morales ». Les expérimentations se feront en étroite collaboration avec l'équipe Facteurs humains de l'ISAE (Institut Supérieur de l'Aéronautique et de l'Espace) - Supaero.

## Contexte

Le travail de thèse se place dans la continuité des travaux sur l'autonomie des systèmes homme-machine, en particulier ceux du Department of Defense américain [8]. Même si ces travaux sont en nombre important, l'apport de la notion d'autorité permet de considérer de manière symétrique l'agent robotique et l'opérateur humain au sein du système. L'originalité des travaux réside dans cette exploitation du concept d'autorité, qui est rarement présent dans la littérature [23, 37, 13] et dans son application aux situations moralement difficiles qui peuvent être créées par l'utilisation de robots armés. De la même manière, nous prenons en compte les forces et faiblesses respectives de chaque agent dans le système, ce qui commence à être fait dans d'autres laboratoires [11].

## Collaborations

La thèse de Florian Gros est co-encadrée par George R. Lucas, Jr., Professor of Ethics & Public Policy, Naval Postgraduate School - Distinguished Chair of Ethics (retired), VADM James Stockdale Center for Ethics, U.S. Naval Academy. Elle est également suivie par Thierry Pichevin, chercheur en éthique aux CREC, écoles de Saint-Cyr Coëtquidan.

Nous participons aux groupes de travail « Régime de responsabilité des robots » et « éthique de la robotique militaire et de surveillance », animés par le CREC.

Catherine Tessier est membre invité de la CERNA, groupe Robotique.

Des collaborations nationales vont être effectives dans le cadre du projet ETHICAA financé par l'ANR (voir paragraphe).

## Formation doctorale

L'école doctorale BSB (Biologie, Santé, Biotechnologies) du PRES Toulouse met en place depuis plusieurs années une formation relative aux « Aspects éthiques généraux de la recherche et enjeux de société ». L'école

doctorale EDSYS (Systèmes, qui recouvre des problématiques de Robotique et d'IA) du site toulousain s'est jointe à l'organisation de cette formation en 2012-2013.

En outre, l'école doctorale EDSYS a mis en place en 2012-2013 un module de formation « Systèmes autonomes et éthique ». Ce module sera reconduit et étendu dans les années à venir, en particulier grâce au cadre du projet ETHICAA (voir paragraphe).

## Ardans Labs

L'ingénierie de la connaissance et la mémoire collective au cœur de la dynamique éthique des organisations

- Alain Berger - [aberger@ardans.fr](mailto:aberger@ardans.fr)
- Ardans SAS
- 6 rue Jean-Pierre Timbaud, "Le Campus" Bâtiment B1,
- 78180 Montigny-le-Bretonneux - France
- <http://www.ardans.fr> - Twitter : @ardanssas
- +33.1.39.30.99.00

### Membres de l'équipe

- Alain Berger, Directeur Général
- Jean-Pierre Cotton, Président et Directeur Technique
- Aline Belloni, Consultant Ingénierie de la connaissance
- François Vexler, Consultant Ingénierie de la connaissance
- Philippe Ducôté, Consultant Ingénierie de la connaissance
- Vincent Besson, Consultant Ingénierie de la connaissance

### Thème général de l'équipe

Ardans a été créée en 1999 et sa vocation est d'aider les entreprises à valoriser leur savoir-faire au sein de leur système d'information. La conviction de la société est bâtie sur le constat que le **savoir-faire constitue l'identité de l'entreprise, sa capitalisation est la valeur ajoutée de son futur**. Aussi, l'équipe travaille sur les trois grandes composantes des opérations suivantes :

1. Structurer les connaissances de l'entreprise et son savoir-faire afin d'en assurer la transmission et la pérennisation,
2. Se fonder sur l'outil Ardans Knowledge Maker<sup>1</sup> moteur conçu et dédié pour valoriser les connaissances,
3. Enrichir simultanément les systèmes d'information et les bases de connaissances.

Pour une description de cette activité on se reportera à l'article *Construire une mémoire collective de l'entreprise : la gestion des connaissances* [7].

### Description des travaux

*L'ingénierie de la connaissance : l'outil de la mémoire collective*

Ardans pratique l'ingénierie de la connaissance principalement dans les classes de situations organisationnelles suivantes : une expertise individuelle, une expertise d'un service ou des services d'un département d'une entreprise, une expertise distribuée dans un réseau d'entités qui appartiennent à différentes organisations.

On définit par **ingénierie de la connaissance**, ce métier de l'Intelligence Artificielle qui est né avec la technologie des systèmes experts il y a plus de quarante ans lors de la rédaction des règles et des premières bases de connaissances *déclaratives*. Cette discipline aujourd'hui couvre tout un cycle depuis l'émergence d'un élément de connaissance, sa détection, sa structuration, son mûrissement en termes de contenu, son expression (*via* une définition claire, non ambiguë, la rédaction étant appuyée par des illustrations si nécessaire), son applicabilité (en termes de domaine d'usage, de droit à en connaître en termes de publication ou de d'habilitation, de durée de vie ou de péremption), et bien sûr de validation (appréciation d'expert, justification, degré de preuve). Lorsque l'on travaille sur la mémoire collective d'un domaine métier, il convient d'orchestrer les différents modèles qui porteront les éléments de connaissances, de poser les liens entre les éléments en relation, et de classifier ces éléments par rapport à des concepts ou des ontologies intelligibles, distinctes, complémentaires et non contestables.

### *Réussir l'implantation d'une mémoire collective*

Une des questions récurrentes concerne les clés du succès d'une telle action de mise en place d'une mémoire collective. L'expérience des opérations conduites depuis

1. Il s'agit d'une plate-forme logicielle basée sur notre expérience en ingénierie de la connaissance et implantant notre méthode [28] et qui s'enrichit par les retours d'expérience des opérations successives.

plus de dix années est que les regards qui sont portés par les informaticiens sont rivés vers la nature de la plate-forme technique, ceux des financiers vers les coûts générés et le montant du retour sur investissement du dispositif, ceux des managers vers la dynamique collective engendrée, la maîtrise du processus métier et la performance collective pérennisées, ceux des ingénieurs de la connaissance sur la fidélité de la restitution du patrimoine de connaissance exploré, la qualité de la structuration pour l'exploitation des éléments de connaissance en situation de travail, la validité du corpus d'expertise colligé par un contenu approuvé par l'expert (ou les experts) et la lisibilité du patrimoine constitué pour une bonne appropriation et une confiance assurée pour l'utilisateur de cette mémoire. Beaucoup de ces points sont des pistes de recherche techniques sur lesquels des résultats d'expérimentation ont été l'objet de publication<sup>2</sup>.

La question qui subsiste est pourquoi dans certains contextes, malgré les faibles moyens ou la difficulté technique, les opérations réussissent alors que dans des cas *a priori* plus favorables c'est l'échec ?

#### *Mémoire collective et éthique*

Pour implanter une mémoire collective dans une entreprise, nous aimons paraphraser Winston Churchill<sup>3</sup> qui n'avait rien à offrir que du sang, du labeur, des larmes et de la sueur. Pourquoi cette mémoire collective prend-elle dans certains cas ? Si la pertinence d'une méthode d'ingénierie des connaissances ou l'adéquation d'un outil logiciel comme plate-forme de travail sont d'une grande importance, comment qualifier : l'envie des acteurs (des managers, experts, gestionnaires, à l'utilisateur lambda), le respect entre eux, une vision partagée ? Ce traitement des principes régulateurs de l'action et de la conduite morale pour une opération... cette dimension subtile et indéterminée correspond parfaitement à ce que l'on définit par le mot *éthique*.

#### **Exemple de projets de mémoire collective intégrant une éthique**

Pour cette clé nommée *éthique* qui correspond à ce climat humain, à cette philosophie qui transforme en terrain de succès ces opérations bien qu'elles soient pointues en termes de méthode, de technique mais qui *in fine* restent essentiellement des aventures humaines, nous avons effectué le choix de ne pas retenir de grandes organisa-

tions industrielles afin de l'illustrer. Il est vraisemblable que l'éthique constitue un critère qui caractérise la qualité d'un climat humain au sein d'une telle organisation... à un moment précis et pour un périmètre donné.

#### *Exemple : Plate-Forme 3C-R*

Fondé en 2006, le Club 3C R (<http://www.3c-r.com/>) est le réseau pour les ressources biologiques. Sont regroupées sous l'appellation Centres de Ressources Biologiques (CRB) toutes les biobanques, tumorothèques et autre « thèques » qui ont pour mission de gérer des échantillons de qualité et parfaitement annotés, réunis ou non en collection, afin de les mettre à disposition de la communauté scientifique pour faciliter le développement de la recherche tant fondamentale qu'appliquée.

Les objectifs Club 3C-R sont de créer des documents répondant à des attentes des CRB, de partager et échanger les informations utiles, et de travailler sur des sujets d'intérêt des CRB et ce dans un cadre de transparence entre les membres.

La **Plate-Forme 3C-R** a été mise en place comme l'outil dédié pour les adhérents, afin de capitaliser l'information produite ou rendue disponible par le Club, de faciliter la recherche et l'utilisation de tous les documents, de valoriser et partager les connaissances dans les domaines d'intérêt pour les CRB, d'aider les actions des groupes de travail, de permettre un contact aisé et direct entre les membres du Club.

Il est certain qu'avec comme modérateur et animateur du réseau, une experte qui a conduit des missions parlementaires pour la révision des lois bioéthiques, le cadre est favorable à un échange serein entre les acteurs. L'enjeu actuel du dispositif est lié à ce que les membres prennent le relais en termes de contribution. On retrouve ici un frein classique du déploiement d'un tel dispositif lié au phénomène de la prise de confiance d'un acteur jusqu'alors *simple lecteur* vers un acteur *moteur et rédacteur*. Lorsque cette bascule aura été réussie par des premiers membres, les autres seront plus enclins à franchir le pas et une nouvelle phase sera entamée.

2. Selon les problématiques ou domaines, les pointeurs que nous recommandons sont : pour les cahiers de laboratoire le retour d'expérience décrit lors d'IC'2011 [6], pour l'industrie automobile [12], de la santé [9], de l'aéronautique [43] et [42], pour l'industrie des procédés lors de  $\lambda\mu$  2012 [26], pour l'appréciation de la qualité [44].

3. Discours de Winston Churchill, le 13 mai 1940, devant la Chambre des Communes.

*Exemple : Elhom*

L'aventure humaine du projet **Elhom**<sup>4</sup> (<http://www.elhom.com/>) est le fruit d'une alchimie créative autour d'un objet technique multi-fonctions et multi-cultures alliant innovation, expérience industrielle et savoir-faire artisanal. Le détail de la genèse de cette épopée qui a débuté en 2008, est décrit dans un *livre blanc* esthétique, chaleureux et extra ordinaire [27]. L'ADN des fondateurs de ce projet responsable est très rapidement traduit dans un « **Pacte Elhom** » voulu inspirant, pérenne, générateur collaboratif d'idées, de richesses humaines, techniques et économiques.

Mais de quoi parle-t-on ? L'objet physique est une *flight case* qui accompagne les acteurs métiers aux quatre coins de la planète. Elle contient une table surface (*i.e.* un écran tactile de 40 pouces ou 101 cm) sur lequel plusieurs acteurs peuvent coopérer. La *boîte à savoir* embarque un environnement complet de gestion d'informations techniques (géographiques, PIDs, documents, portail internet etc.) configuré pour la mission et qui interagit avec des périphériques mobiles (*smartphones, caméras, etc.*) sans fil comme avec des bases de connaissances métier (**Commissioning** par exemple) sur des serveurs sécurisés *via* des liaisons spécialisées.

La dynamique qui est autour de ce projet est considérable. Elle touche tant les membres qui ont contribué à la naissance des premiers prototypes que les prospects qui ont découvert l'enthousiasme et la technologie associés.

S'il fallait retenir une facette fascinante et enrichissante du poids de l'éthique dans le projet Elhom, ce serait celle liée à la contribution collective et partagée de la mémoire collective des différentes sociétés d'ingénierie qui est formalisée dans les structures (Ontologies et Modèles) partagées. Ce projet cristallise l'idée d'une construction collective d'un langage commun pour bâtir un environnement de coopération sur un domaine métier exhaustif de production industrielle<sup>5</sup>.

*Exemple : PEINTURE chez ACG Industrie*

Après ces deux exemples où des réseaux partagent des connaissances, il est intéressant d'évoquer un cas où une petite entreprise industrielle mise sur le partage des connaissances et plus précisément sur la valorisation de sa mémoire collective de recherche et développement afin

de gagner en performance pour sa pérennité et son développement.

ACG Industrie est une entreprise dont l'activité originelle au sens de l'INSEE est dédiée à la « Fabrication de peintures, vernis, encres et mastics ». Elle a été créée en 1966 et suite à des périodes économiques difficiles, elle a été reprise début 2012 par un industriel (EuroFIP) qui a chargé son nouveau Directeur Général de redonner un nouveau souffle à ce site de 30 salariés : défi sérieux si l'on considère que l'effectif avait été réduit de moitié dans les deux précédentes années.

C'est donc le Laboratoire qui a été la source du projet de mémoire collective : PEINTURE<sup>6</sup>. La question qui est soumise quotidiennement aux chercheurs est celle d'élaborer la bonne formulation, la bonne recette pour que le produit *in fine* satisfasse le client. Les étapes de validation autour de la paillasse sont multiples : l'application sur le matériau, la qualité face aux contraintes de production puis dans la vie du produit (vieillesse en particulier), le rendu en couleur, en aspect et plus encore. Le temps de mise au point du produit, au-delà du *time to market* du client final, est une clé de la rentabilité du processus de production du produit.

Le cadre éthique de l'opération est porté d'un côté sur la volonté collective de redonner à l'entreprise un futur resplendissant, sur la motivation partagée d'être plus efficace collectivement (améliorer les mises au point des produits nouveaux sur la base des retours d'expérience validés et ne pas refaire des manipulations déjà réalisées). L'exemple de l'implication de la Direction dans le dispositif a été un signal fort pour les acteurs du Laboratoire, et la première récompense a été celle des auditeurs pour la certification qui ont apprécié la nouvelle impulsion collective. Ainsi que le rappelle l'actionnaire, « *c'est lorsque de nouveaux contrats seront gagnés en s'appuyant sur l'outil et la démarche que l'on pourra parler de résultat significatif!* ».

Dans la mise en place d'une mémoire collective, si l'éthique se positionne comme un indicateur d'une qualité de vie humaine dans l'entreprise, le marqueur de la performance du dispositif industriel complet reste le rendement économique de l'établissement concerné. La qualité de la mémoire collective est un atout pour la pérennité sans être pour autant un gage absolu.

4. Elhom est la contraction des mots éléphant et homme. L'éléphant symbole de mémoire, de sagesse, de longévité évoque l'éveil et la connaissance, résume une dimension essentielle de la démarche de capitalisation des connaissances et données techniques dans le cycle de vie d'un site industriel ou d'un territoire.

5. Il s'agit du secteur de l'énergie *Oil & Gaz* et "*Électricité*" initialement.

6. PEINTURE est l'acronyme de « *Pérenniser l'Expertise Industrielle et les Nombreuses Techniques Utilisées et Reconnues au sein de l'Entreprise* ».

## Perspectives

Nous sommes dans l'attente de décision sur le financement de projet de recherche avec des acteurs académiques et industriels où l'appréciation de l'éthique d'un système sera traduite en règles. Ces dernières seront stressées par les événements qui feront apparaître des conflits qu'il convient de résoudre de manière justifiée.

La formalisation de ces règles, la description des *scenarii* d'usage, les événements générés et les résolutions de conflits seront autant de connaissances qu'il faudra formaliser et tracer.

Si l'éthique est une clé qui a longtemps été perçue sans porter de nom pour les actions d'ingénierie de la connaissance à des fins de « Mémoire Collective », la formalisation de ces obstacles potentiels devrait apporter de la lisibilité dans la faisabilité des opérations. Par ailleurs, l'usage de technologie d'ingénierie des connaissances pour superviser ou accompagner des systèmes multi-agents dans la gestion des conflits d'éthique est un axe de recherche récent et tout à fait captivant.

## Le projet ANR ETHICAA

Cadre général pour concevoir des systèmes d'agents artificiels dotés de comportements éthiques - Représentation formelle des conflits éthiques et outils de gestion de ces conflits

- CONTACT : Grégory Bonnet - [gregory.bonnet@unicaen.fr](mailto:gregory.bonnet@unicaen.fr)
- ADRESSE : GREYC Université de Caen Basse Normandie

## Partenaires

ETHICAA « ETHICs and Autonomous Agents » est un projet qui a été présenté à l'Agence Nationale de la Recherche (ANR) lors de l'appel à projet CONTINT 2013 (Contenus numériques et interactions) et accepté pour une durée de 4 ans (2014-2017). Il est porté par le consortium GREYC Université de Caen Basse Normandie (coordonnateur du projet), ARMINES - FAYOL (Institut Henri Fayol de l'école des Mines de Saint-étienne), l'ONERA-DCSD (voir paragraphe), le LIP6 (Laboratoire d'Informatique de Paris 6) (voir paragraphe), l'équipe de recherche ETOS de Télécom école de Management (Institut Mines-Télécom) et Ardans (voir paragraphe).

## Contexte

À l'heure actuelle, les machines et les agents artificiels deviennent de plus en plus autonomes et, par conséquent, agissent de plus en plus sans être contrôlés par des utilisateurs ou des opérateurs humains. Toutefois, et en particulier lorsque les machines autonomes interagissent directement avec des humains, il est de première importance de s'assurer qu'elles ne peuvent les blesser, ni menacer leur propre autonomie. C'est pourquoi la question de doter les agents autonomes de comportements éthiques régulés ou contrôlés se pose.

L'objectif du projet ETHICAA est de définir ce que devrait être une entité autonome pouvant gérer des conflits éthiques, aussi bien au niveau individuel qu'au niveau collectif. Les conflits éthiques que nous considérons sont susceptibles de survenir au sein d'un agent, entre un agent et les règles éthiques de son système d'appartenance, entre un agent et un utilisateur ou un opérateur humain, entre plusieurs agents incluant ou non des êtres humains. Les conflits éthiques sont caractérisés par le fait qu'il n'existe pas de « bonne » manière de les résoudre. Quoi qu'il en soit, lorsqu'une décision doit être prise, ce devrait être une décision informée se fondant sur l'évaluation des arguments qui la soutiennent et des conséquences qu'elle produit. Le projet ETHICAA, en produisant des modèles, des méthodes et des outils, proposera un cadre méthodologique pour définir et gérer des agents autonomes dotés de comportements éthiques.

## Domaines d'application

Les deux domaines d'application proposés par ETHICAA sont la robotique et la gestion de la vie privée. Ils ont été retenus pour les questions de société qu'ils soulèvent en termes de confiance, d'acceptation et d'éthique. Dans le domaine de la robotique, le cadre ETHICAA sera utilisé pour gérer les conflits éthiques qui surviendraient lors de la coopération entre un agent artificiel et un opérateur humain dans le contexte du contrôle conjoint d'un drone, le système homme-machine étant soumis à des perturbations et à des dilemmes éthiques.

Dans le domaine de gestion de la vie privée, les conflits éthiques peuvent survenir entre plusieurs agents artificiels et plusieurs utilisateurs humains. Le cas d'un réseau social dans lequel des agents éthiques artificiels contrôlent et gèrent les politiques de vie privée de leurs utilisateurs humains sera considéré.

## Résultats attendus

À partir de la définition et l'expérimentation de ces scénarios, le projet ETHICAA a pour objectif de fournir une représentation formelle des conflits éthiques et des objets sur lesquels ils portent, ainsi que des outils pour pouvoir les gérer. En conséquence, le résultat attendu du projet ETHICAA est un cadre général pour concevoir des systèmes d'agents artificiels dotés de comportements éthiques, c'est-à-dire de concevoir les mécanismes de contrôle et de régulation de fonctions autonomes devant obéir à des règles éthiques dépendantes d'un contexte et devant tenir compte des conflits éthiques impliquant d'autres agents artificiels et des êtres humains, quel que soit leur comportement.

## Le CERNI

Comité d'éthique pour les Recherches non Interventionnelles

- CONTACT : [comite-cerni@grenoblecognition.fr](mailto:comite-cerni@grenoblecognition.fr)
- WEB : <http://www.grenoblecognition.fr>

## Contexte

Le Pôle Grenoble Cognition est une Structure Fédérative de Recherche (FR 3381) du CNRS (INSII) et de l'université de Grenoble. Il se compose de scientifiques et d'équipes de recherche en science cognitive issus de 15 laboratoires grenoblois. Son rôle est de favoriser les collaborations et de faciliter la progression des connaissances dans un domaine très complexe, aux enjeux sociétaux importants pour l'éducation, la santé, le handicap, les télécommunications ... Les recherches en science cognitive impliquant régulièrement des expérimentations comportant des sujets humains, il est apparu important au sein du pôle d'aborder les questions relatives à l'éthique. Sous l'impulsion d'un groupe de volontaires, il a été décidé de créer un groupe de travail pour la mise en place d'un comité d'éthique.

## Création du comité

En étroite collaboration avec le Groupe de Recherche « Droit et Science » de l'Université Pierre-Mendès-France, ce groupe de travail a rédigé les statuts et défini les grandes lignes d'une charte – en cours de finalisation –

7. <http://www.grenoblecognition.fr>

et les règles de fonctionnement. Une fois ces règles fixées, une démarche d'information a été initiée auprès des présidents des établissements d'enseignement supérieur et recherche de Grenoble afin de solliciter leur appui dans la création officielle du comité d'éthique : le CERNI.

Ce comité d'éthique a pour fonction de donner un avis sur des protocoles de recherche impliquant la personne humaine qui ne relèvent pas de la recherche biomédicale et ne sont donc pas actuellement soumises au contrôle d'un Comité de Protection des Personnes (CPP). Dans certaines disciplines, la publication dans les grandes revues internationales exige qu'un avis favorable d'un comité d'éthique accompagne la demande de publication d'un article. De plus en plus d'agences de moyens conditionnent l'octroi de financements à l'obtention d'un avis favorable d'un comité d'éthique (PCRDT, financements internationaux et bientôt ANR). Enfin, une proposition de loi, en cours d'examen au Parlement à l'époque de la mise en place de ce comité et qui a été adoptée depuis, impose de soumettre tous les protocoles de recherche sur la personne humaine à un comité d'éthique (loi Jardé).

En dehors des aspects plus administratifs, ce comité d'éthique se veut un lieu de réflexion. Il doit pouvoir être facilement saisi par n'importe quelle équipe de recherche du pôle souhaitant obtenir un avis sur sa démarche scientifique.

## Mode de fonctionnement

Une équipe souhaitant solliciter le comité dans le cadre d'un dépôt de projet doit remplir un dossier de présentation, exposant les différents points spécifiés dans un guide mis en ligne<sup>7</sup>. Ce dossier doit en particulier préciser de manière non exhaustive les objectifs du projet, le mode de recrutement des participants, le respect de la confidentialité des données, les conditions d'archivage des résultats, la notice d'information et de consentement éclairé ainsi qu'une évaluation du rapport bénéfice / risque encouru. Un tableau de critères est proposé pour faciliter l'évaluation de ce dernier point (duperie des sujets, atteinte à la vie privée, à l'estime personnelle, ...) Une fois la commission saisie, deux rapporteurs sont désignés pour l'étudier. Le projet est ensuite présenté (une séance tous les deux mois). Le comité peut rendre cinq catégories d'avis : favorable, défavorable, favorable sous réserve de modifications mineures, modifications majeures à soumettre au CERNI, requalification vers le CPP. à l'exception de l'avis favorable, les décisions du CERNI s'accompagnent de conseils personnalisés pour améliorer le dossier dans ses aspects éthiques.

Depuis sa création il y a un an et demi, le comité s'est régulièrement réuni tous les deux mois et étudie un nombre croissant de dossiers.

## Comptes rendus de workshops

Nous donnons ci-dessous les compte rendus de deux ateliers qui ont eu lieu en 2012 sur les questions de l'éthique et de l'autonomie d'agents.

### Atelier « Agents Autonomes et éthique » à RFIA 2012

Quelles questions éthiques liées aux capacités informatiques conférant de plus en plus d'autonomie aux logiciels, aux robots ?

- CONTACT : Olivier Boissier, Catherine Tessier
- WEB :

<http://d2a2.emse.fr/agents-autonomes-ethique-rfia/>

L'atelier a réuni 13 participants le 24 janvier 2012 à Lyon, dans le cadre des événements associés à RFIA 2012.

Au travers de 6 exposés, cet atelier a permis de partager différentes visions et approches sur la question de l'éthique dans le cadre général des STIC et plus particulièrement de celui des agents ou systèmes autonomes.

Dans son exposé, Joseph Mariani (LIMSI-CNRS) a présenté la synthèse du rapport élaboré dans le cadre du COMETS (COMité d'ETHique du CNRS) qui s'est intéressé à l'éthique de la pratique des recherches dans le domaine des STIC en l'élargissant également à l'éthique des technologies et des applications. Ce travail a ainsi permis de recenser différentes questions relatives à : la protection de la vie privée, l'impact des STIC sur l'homme, le robot comme une personne, les risques sanitaires technologiques, le travail, l'accès au savoir et à la formation, les échanges économiques, la vie collective. Ce travail montre que, pour être plus éthique, la recherche en STIC nécessiterait d'être accompagnée d'une réflexion sur les conséquences de ses résultats. Cependant en France, peu d'actions sur le long terme sont menées, et l'attitude des chercheurs semble frileuse par rapport à ces questions. La participation française dans des comités et travaux sur l'éthique à l'international est faible.

De cette étude, ressortent 8 recommandations : créer au niveau national un comité d'éthique STIC, commun aux

organismes de recherche nationaux et pluridisciplinaire (N.B. : depuis la CERNA a été mise en place au sein d'Allistène), soutenir des projets de recherche communs STIC et SHS, encourager les recherches en STIC dans des domaines clefs relatifs aux questions d'éthique, faciliter l'accès aux données, identifier les incidences éthiques des recherches, préférer l'information objective sur l'avancée des recherches en STIC plutôt que les effets d'annonces, former à l'éthique des STIC, et enfin, sensibiliser les chercheurs et les citoyens.

Le deuxième exposé donné par Olivier Vasseur (ONERA), s'est intéressé aux perceptions de l'éthique. Il a mis en lumière la variabilité des principes éthiques, la diversité intrinsèque et les divergences d'appréciation. Certains affirment que les comités d'éthique sont une entrave à la liberté de faire de la recherche (Michael Burawoy, Université de Californie-Berkeley, *in* congrès AFS 2009). Les attentes entre académiques et non-académiques sont différentes notamment en ce qui concerne l'utilité d'une charte éthique (le privé serait « moins éthique » que le public). Cette diversité apparaît également dans l'étude réalisée par l'Unesco de 65 codes de conduite dans différents pays : des différences existent dans les approches et donc dans les pratiques (par exemple, usage variable du consentement éclairé selon les pays). Il peut donc y avoir des conflits entre les directives. Les diversités culturelles et l'absence de référence supranationale font également qu'il ne peut pas y avoir de solution type. L'éthique peut ainsi apparaître comme un frein à la recherche, sachant que l'éthique ne peut pas résoudre les problèmes de droit et qu'inversement les problèmes d'éthique ne peuvent pas tous être résolus par le droit. Terminant sa présentation en se penchant plus particulièrement sur les robots autonomes, Olivier Vasseur pose différentes questions : quel est le niveau de décision qui est délégué aux agents, quelle est l'influence des valeurs personnelles ou religieuses ? quel équilibre adopte-t-on entre différentes perceptions éthiques ? quelle relation d'autorité avec l'utilisateur (par exemple : se faire euthanasier par son robot ?) Le robot serait-il un moyen d'uniformisation de l'éthique ?

La présentation de Thierry Pichevin (écoles de Saint-Cyr Coëtquidan), quant à elle, traite des questions posées par l'utilisation d'agents autonomes dans le domaine militaire (traitement des mines, renseignement, surveillance, allègement du combattant ; robots combattants, à capacité de destruction, drones armés). En préambule, T. Pichevin rappelle que les robots doivent s'inscrire dans la théorie de la guerre juste (*Jus in bello*)

et respecter les principes de discrimination (distinguer combattant et non-combattant, y compris les installations) et de proportionnalité (le bénéfice doit être supérieur aux dégâts). Deux courants de pensée s'opposent parmi les chercheurs en robotique. Ainsi, Noël Sharkey (University of Sheffield) affirme qu'un robot ne peut pas respecter les deux principes : en ce qui concerne la discrimination, les textes sont flous, le bon sens est requis, les machines ont du mal à discriminer des objets d'aspects proches ; pour la proportionnalité, il n'y a pas de méthode objective, elle repose sur l'intuition. En revanche, Ron Arkin (Georgia Tech), qui met en œuvre les lois de la guerre dans les robots, rappelle que les humains violent les critères du *Jus in bello* (perte d'objectivité par peur, vengeance, haine, colère) et que donc des machines, qui n'ont pas ces émotions, devraient pouvoir faire mieux : elles perçoivent mieux, elles sont plus rapides, plus endurantes, moins vulnérables aux armes chimiques, ne subissent pas de conséquences psychologiques et appliquent des critères froids et objectifs. Sur ces deux discussions, T. Pichevin fait remarquer que la plupart des arguments sont très techniques, utilitaristes, et considèrent essentiellement les conséquences. Il rappelle que l'on ne peut pas s'arrêter là : par exemple, comment traiter les cas où les règles se contredisent ? Un équilibre est à trouver entre machine et homme, la machine étant en appui. Les réflexions éthiques (c'est-à-dire, l'art de régler les conflits de valeurs) sont à mener au fur et à mesure plutôt qu'après coup.

Danièle Bourcier (CERSA-CNRS) aborde cette question de l'éthique dans le domaine des lois et de la justice. Elle commence son exposé en montrant comment l'IA permet de simuler, de prévoir les effets d'un système de normes. Les fonctions cognitives les plus nobles (décider, juger, légiférer) peuvent donc être assistées ou simulées par des machines (exemple : radar automatique). Cette vision rentre en conflit avec celle développée par Joseph Weizenbraum (MIT) qui disait qu'on ne peut éthiquement automatiser ni le droit ni la psychiatrie. Cette vision s'appuie sur la question de la manière de représenter la notion de marge d'appréciation dans un processus décisionnel (« se conduire en bon père de famille », intuition, sens commun). Est-ce que les machines ont la capacité technique, juridique, éthique de décider ? Une réponse normative est donnée, par exemple, dans les articles 2 et 3 de la loi de 1978 qui stipulent qu'une décision de justice ne peut être fondée sur un traitement automatisé ; qu'une décision administrative ou privée ne peut être fondée seulement sur un traitement automatisé. D. Bourcier affirme ainsi qu'une réponse éthique repose sur les principes de précaution, de compétence,

de responsabilité (capacité pour un sujet/agent compétent et éclairé de produire un acte de façon autonome et d'assumer les conséquences de cet acte), de gouvernance. Un ensemble de questions doivent être posées quant à l'utilisation de systèmes autonomes dans le domaine du droit : a-t-on le droit de fabriquer certains outils ? a-t-on le droit de les utiliser, notamment face à une population qui va en subir les impacts ? le droit doit-il être réduit à un système de gestion dans un cadre de décision autonome prise par des machines ?

D. Bourcier termine ainsi sa présentation en rappelant que la machine ne doit pas régler plus de problèmes que n'en sait résoudre l'homme. Il faut considérer la boucle homme-machine, sans opposer les deux types d'entités. Les questions d'éthique ne peuvent pas être réglées dans des machines.

Après ces exposés posant les questions globales de l'éthique en terme d'usage et de domaines d'application, les deux exposés suivants abordent les questions de modèles et de formalisation.

Jean-Gabriel Ganascia (LIP6-CNRS) pose la question de la modélisation et de la résolution des conflits éthiques à l'aide de logiques non monotones. L'éthique réfléchit aux fondements des systèmes de règles et sur la résolution des conflits de normes. C'est la nature même de l'éthique que d'avoir des conflits (par exemple les lois d'Asimov sont contradictoires entre elles). Même si Ron Arkin suppose que les systèmes de règles peuvent être facilement implémentés dans les robots, ce n'est pas le cas.

On peut envisager des formalisations par la théorie des jeux, les logiques déontiques (qui peuvent aboutir à des instances du paradoxe de Chisholm : obligation morale de faire ce qui est interdit), le raisonnement à partir de cas, l'apprentissage statistique, les logiques non monotones. Dans ce dernier cadre, on modélise le fait que l'ajout d'information va invalider des règles. Si on considère des agents BDI, Belief (B) va contenir les perceptions, les valeurs (relation d'ordre), l'évaluation de conséquences des Intentions (I). Dans le cas de conflits éthiques, il s'agit de choisir l'intention dont les conséquences sont les moins mauvaises : c'est l'échelle de valeurs qui permet de lever le conflit.

Dans sa présentation, Guillaume Piolle (SUPELEC) s'intéresse à un agent devant veiller à la protection de la vie privée de son propriétaire et d'autres personnes et aux règles relatives à la collecte, la conservation, l'utilisation, la diffusion d'informations se rapportant à la vie

privée : dans ce contexte, l'agent va-t-il rester un agent, être autonome ?

La mise en œuvre des normes (textes internationaux, lois nationales, textes réglementaires, contrats, réglementations internes, directives spécifiques, préférences utilisateur) suppose de passer d'un texte à un langage formel et de modéliser les modalités : interdiction, obligation, permission (qui peuvent être des exceptions aux interdictions) ainsi que les notions temporelles : obligation avec échéance, obligation de présence, interdictions maintenues. Comment combiner les modalités déontiques et temporelles ?

Il faut s'assurer de la cohérence des normes, spécifier quelles normes sont à utiliser dans quel contexte, contraindre le comportement de l'agent selon les normes actives, effectuer des calculs de risque sur le comportement d'autres agents. En ce qui concerne les conflits, détecte-t-on les conflits potentiels ou réels (surtout dans le contexte déontico-temporel), fait-on le calcul en ligne ou hors ligne ?

Doit-on concevoir des agents normatifs ou des agents éthiques ?

### RDA2, ECAI Workshop on Rights and Duties of Autonomous Agents

Autonomie des agents, partage de l'autorité : point de vue formel, point de vue moral ou légal

- CONTACT : Grégory Bonnet (compte rendu), Olivier Boissier, Catherine Tessier
- WEB : <http://rda2-2012.greyc.fr/>

L'atelier RDA2 a eu lieu le 28 août 2012 à Montpellier, dans le cadre de la conférence ECAI 2012. Il avait pour objectif de regrouper différentes personnes issues des sciences humaines et sociales, de la philosophie, de l'informatique et de la robotique autour des questions relative à l'autonomie des agents et au partage de l'autorité aussi bien d'un point de vue formel que d'un point de vue moral ou légal. 10 articles ont été soumis à l'atelier, et 7 ont été acceptés, chaque article ayant été évalué par trois relecteurs. Une vingtaine de personnes a assisté aux présentations et discussions. Un exposé invité a été donné par A.C. Da Rocha Costa (FURGRS-Brésil).

Les actes de l'atelier sont disponibles sur CEUR (<http://ceur-ws.org/Vol1-885/>).

Au regard de ce qu'ont présenté les orateurs, trois angles d'attaque permettent d'aborder la question : le

raisonnement formel, la réflexion philosophique ou l'application.

### Exposés

D'un point de vue formel, Antônio Carlos da Rocha Costa (**Functional Rights and Duties at the Micro and Macro Social Levels**) distingue clairement les droits et devoirs des agents autonomes des permissions et obligations que ceux-ci doivent satisfaire. En effet, les premiers concernent les interactions, c'est-à-dire un lien entre agents, tandis que les seconds concernent les comportements, c'est-à-dire les actions d'un unique agent. L'idée principale est que les droits et devoirs sont structures d'interaction entre les rôles des agents. Cela définit seulement un cadre possible pour les interactions, car le droit ne spécifie en rien le devoir associé. Antônio Carlos da Rocha Costa propose une formalisation dans une logique modale : les droits et devoirs sont des opérateurs ternaires, entre deux agents autour d'une action. Cette notion est à rapprocher de celle d'engagement. On peut aussi avoir une représentation graphique de ces interactions, qui sont à rapprocher cette fois des travaux de Mark Burgess sur la théorie de la promesse<sup>8</sup> dans les réseaux.

Mihnea Tufiş (**Normative Rational Agents - A BDI Approach**) se positionne sur une approche plus traditionnelle et propose une architecture d'agent BDI normatif. Il y a deux bases de normes, une abstraite, c'est-à-dire où les normes ne sont pas encore acceptées, et une instanciée qui contient toutes les normes actives valides. La consistance est donnée par l'ensemble maximal des normes sans conflit dont les conséquences sont les moins néfastes au sens des plans que peut formuler l'agent. Il est intéressant de noter que cela semble très proche d'un raisonnement d'argumentation formelle.

D'un point de vue philosophique, Florian Gros (**Ethics and Authority Sharing for Autonomous Armed Robots**) se demande si l'on désire réellement des robots complètement autonomes. Si oui, il faut se poser la question de la responsabilité. Si Kant voit la responsabilité comme transcendante (notion d'impératif catégorique), Florian Gros suit plutôt une pensée immanente à la manière d'un contrat social rousseauiste. Trois voies semblent possibles : une responsabilité réduite de la machine au profit du constructeur, une responsabilité partielle au profit d'un de ses composants, ou lui donner un statut moral, la machine étant traitée *comme* un humain.

8. Voir <http://research.iu.hio.no/promises.php>

Alberto Pirni (**Subjectivity of Autonomous Agents. Some Philosophical and Legal Remarks**) distingue toutefois clairement la question philosophique de la question légale. Il passe en revue un certain nombre de théories sur l'autonomie et la conscience, et conclut qu'à l'heure actuelle aucun robot ne peut être qualifié d'autonome au sens philosophique du terme. Il fait le même travail d'un point de vue légal, et en ressort deux possibilités : (1) le robot peut être considéré selon la législation de l'animal (dans la droite ligne d'un Morguénaud ou un Feinberg) ; (2) les agents négociant peuvent être considérés comme des personnes mineures, c'est-à-dire responsables mais limités à de petites transactions.

Bart Kamphorst (**The Primacy of Human Autonomy : Understanding Agent Rights Through the Human Rights Framework**) part quant à lui du postulat qu'un agent artificiel<sup>9</sup> préserve l'autonomie de l'humain en termes de choix de vie ainsi que sa dignité telle que la définit (ou plutôt ne la définit pas) la Déclaration universelle des droits de l'homme. Il propose alors quatre lignes directrices devant sous-tendre toute interaction entre un humain et un agent autonome : l'agent doit (1) préserver la dignité ; (2) préserver l'autonomie à moins d'y être forcé par la loi ; (3) ne déléguer une tâche que dans le but de satisfaire un but ; (4) la délégation doit toujours préserver l'autonomie.

D'un point de vue applicatif, ce sont essentiellement les drones autonomes et leurs opérateurs humains qui sont considérés.

Thomas Dubot (**Integrating Civil Unmanned Aircraft Operating Autonomously in Non-Segregated Airspace : Towards a Dronoethics ?**) s'intéresse essentiellement aux drones civils. Il y a beaucoup d'applications pour les engins volants autonomes mais l'espace aérien est aujourd'hui trop restreint en termes d'opération et de régulation pour leur déploiement à grande échelle. Thomas Dubot tente alors de faire une taxonomie des enjeux et des types de décisions qui peuvent être prises dans un contexte homme-machine.

Sergio Pizziol (**What the Heck Is It Doing? Better Understanding Human-Machine Conflicts Through Models**) propose de modéliser les conflits dans l'interaction homme-machine en vue de les détecter et de les résoudre. La question qui revient dans ces travaux est celle de la fiabilité de l'interaction entre l'humain et la machine et du comportement de la machine qui, selon la législation et les procédures, doit prendre les mêmes décisions que celles que pourrait prendre un humain dans la même situation.

#### Discussion

Suite aux présentations des articles, une session de discussion a eu lieu. Il avait été demandé à chaque orateur de relire l'article d'un autre orateur, et présenter en quelques minutes les réflexions qui lui étaient venues à ce sujet. Ce mode de fonctionnement a été fructueux et a pu ouvrir de nombreuses discussions. Il en ressort :

1. Qu'il n'est peut-être pas pertinent de parler d'autonomie en soi. Ce terme est difficile à définir et dépend du contexte. Il convient d'utiliser plutôt le terme d'autorité en pensant un cadre dans lequel il faut décider quel agent a de l'autorité sur une ressource pour l'utiliser à sa convenance. Il devient alors intéressant de se poser la question de la responsabilité quant à l'utilisation de ces ressources.
2. Qu'il faut peut-être remettre en cause l'hypothèse qui donne toujours la priorité à l'humain, ce dernier pouvant être défaillant. Cela est à mettre en perspective avec le fait que la loi cherche avant tout à savoir qui est responsable de quoi. D'un point de vue application, il convient de noter qu'il existe déjà des systèmes à bord des avions qui limitent certaines actions de l'humain. Qui est donc responsable des actions de ces systèmes ?
3. Qu'il faut peut-être dépasser les dialectiques traditionnelles comme celle du producteur-consommateur ou celle du maître-esclave, et qu'il faut mettre en question notre biais anthropocentrique. En ce sens, les modélisations à l'aide d'états mentaux sont-elles les plus à même de capturer toutes ces notions ?

Pour finir, il convient peut-être de décomposer la question initiale en une série de questions répondant au schéma : quels sont les  $A$  d'un  $X$  au regard des  $B$  d'un  $Y$  ? où  $\{A, B\}$  et  $\{X, Y\}$  sont respectivement toutes les combinaisons de {droits, devoirs} et {agent artificiel, humain}.

9. L'auteur s'intéresse aux applications d'e-Coaching où un agent artificiel conseille un humain sur son style de vie.

## Bibliographie

- [1] N. Abchiche-Mimouni, A. Andriatrimoson, E. Colle, and S. Galerne. Multidimensional adaptiveness in multi-agent systems. *International Journal on Advances in Intelligent Systems*, 6(1-2) :124–135, July 2013.
- [2] A. Andriatrimoson, N. Abchiche-Mimouni, E. Colle, and S. Galerne. An adaptive multi-agent system for ambient assisted living. In *ADAPTIVE 2012*, pages 85–92. IARIA, ThinkMind, jul 2012.
- [3] M. Aupetit. Visualizing distortions and recovering topology in continuous projection techniques. *Neurocomputing, Elsevier*, 70 :1304–1330, 2007.
- [4] M. Aupetit, L. Allano, I. Espagnon, and G. Sannie. Visual analytics to check marine containers in the eritr@c project. *Proc. of the International Symposium on Visual Analytics Science and Technology (EuroVAST), Bordeaux*.
- [5] Anthony Beavers. *Robot Ethics : The Ethical and Social Implication on Robotics.*, chapter Moral Machines and the Threat of Ethical Nihilism. MIT Press, Cambridge, Mass., January 2012.
- [6] A. Belloni, A. Berger, JP. Cotton, and F. Devoret. De la gestion des connaissances structurées au cahier de laboratoire électronique à valeur probatoire : naissance de CLEOPATRE ([www.ardans.fr/doc/IC2011\\_CLEOPATREpublibook.pdf](http://www.ardans.fr/doc/IC2011_CLEOPATREpublibook.pdf)). , Mai 2011.
- [7] A. Berger and JP. Cotton. Construire une mémoire collective de l'entreprise : la gestion des connaissances ([www.ardans.fr/doc/AFIA\\_Communication.pdf](http://www.ardans.fr/doc/AFIA_Communication.pdf)). *AFIA n° 72*, pages 70–73, Avril 2011.
- [8] Defense Science Board. The role of autonomy in DoD systems. Technical report, Department of Defense, July 2012.
- [9] J. Chanliau, A. Mariot, A. Caillette-Baudoin, P. Mariot, A. Berger, and F. Vexler. L'efficacité des organisations de santé : la bascule dans la gestion des connaissances ([www.irit.fr/SDC2006/cdrom/contributions/Jacqueschaliau\\_SDC2006.pdf](http://www.irit.fr/SDC2006/cdrom/contributions/Jacqueschaliau_SDC2006.pdf)). *ACTES de la SdC 2006*, Juin 2006.
- [10] E. Colle and S. Galerne. Mobile robot localization by multiangulation using set inversion. *Robot. Auton. Syst.*, 61(1) :39–48, January 2013.
- [11] N.J. Cooke and R.A. Chadwick. Lessons learned from human-robotic interactions on the ground and in the air. In M. Barnes and F. Jentsch, editors, *Human-Robot Interaction in Future Military Operations*, pages 355–372. Ashgate, 2010.
- [12] C. Coppens, J. Laroque-Malbert, A. Berger, P. Mariot, and JP. Cotton. La capitalisation des connaissances dans l'industrie : implanter un référentiel métier et le déployer : l'exemple d'Icare ([www.irit.fr/SDC2006/cdrom/contributions/Alainberger\\_SDC2006.pdf](http://www.irit.fr/SDC2006/cdrom/contributions/Alainberger_SDC2006.pdf)). *ACTES de la SdC 2006*, Juin 2006.
- [13] M.L. Cummings and J.C. Ryan. Shared authority concerns in automated driving application. MIT Humans and Automation Laboratory Working Paper, 2013.
- [14] Anne Fagot-Largeault. Petites et grandes fraudes scientifiques . la mondialisation de la recherche. *Conférence du collège de France*, 49(4) :1417 – 1430, 2011.
- [15] P. Gaillard, M. Aupetit, and G. Govaert. Learning topology of a labeled data set with the supervised generative gaussian graph. *Neurocomputing*, 71(7-9) :1283–1299, 2008.
- [16] J.-G. Ganascia. Modelling ethical rules of lying with answer set programming. *Ethics and Information Technology*, 9(1) :39–47, 2007.
- [17] J.-G. Ganascia. The new ethical trilemma : Security, privacy and transparency. *Comptes Rendus Physique*, 12 :684–692, September 2011.
- [18] F. Gros, T. Pichevin, E. Pomes, and C. Tessier. L'« autonomie » de la machine dans les systèmes homme-machine : évolution ou révolution du champ de bataille? aspects juridiques et éthiques. *Dynamiques Internationales*, 2013. À paraître.
- [19] F. Gros, C. Tessier, and Th. Pichevin. Ethics and authority sharing for autonomous armed robots. In *CEUR Workshop Proceedings*, editor, *RDA2, Proceedings of the ECAI 2012 Workshop on Rights and Duties of Autonomous Agents*, volume 885, Montpellier, France, 2012.
- [20] J. Hansen. The paradoxes of deontic logic. *Theoria*, (72) :221–232, 2006.
- [21] J. Horty. Moral dilemmas and nonmonotonic logic. *Journal of Philosophical Logic*, (23) :35–65, 1994.
- [22] J. Horty. *Defeasible Deontic Logic*, chapter Nonmonotonic foundations for deontic logic, pages 17–44. Kluwer Academic Publishers, 1997.

- [23] T. Inagaki. Automation and the cost of authority. *International Journal of Industrial Ergonomics*, 31 :169–174, 2003.
- [24] J.-P. Poli J.-M. Le Yaouanc. A fuzzy spatio-temporal-based approach for activity recognition. *ER'12 Proceedings of the 2012 international conference on Advances in Conceptual Modeling, Heidelberg*, 2012.
- [25] S. Lespinats and M. Aupetit. Checkviz : sanity check and topological clues for linear and nonlinear mappings. *Computer Graphics Forum*, 30(1), 2011.
- [26] A. Louis. ALPIN, une application collaborative de partage de connaissances, autour des documents techniques clés de l'ingénierie d'AIR LIQUIDE Engineering ([www.imdr.fr/docs/LM18\\_Actes%20congres/textes/lm18\\_com\\_3C-5\\_079\\_A\\_Louis.pdf](http://www.imdr.fr/docs/LM18_Actes%20congres/textes/lm18_com_3C-5_079_A_Louis.pdf)). *Actes Congrès 18<sup>ème</sup>λμ*, Octobre 2012.
- [27] Y. Mahé, L. Ramolino, F. Le Rolland, C. Soriano, F. Barbin, A. Hentinger, F. Renaud, F. Clermont, and M. Félez. On the road with elhom. Un autre monde est en marche, Dexip (<http://www.elhom.com/>), 2012. Livre Blanc.
- [28] P. Mariot, C. Golbreich, , JP. Cotton, and A. Berger. Méthode, Modèle et Outil Ardans de capitalisation des connaissances ([www.med.univ-rennes1.fr/lim/doc\\_166.pdf](http://www.med.univ-rennes1.fr/lim/doc_166.pdf)). *RNTI EGC 2007*, pages 187–206, Janvier 2007.
- [29] D. Mercier, P. Gaillard, M. Aupetit, C. Maillard, R. Quach, and J.-D. Muller. How to help seismic analysts to verify the french seismic bulletin? *Engineering Applications of Artificial Intelligence*, 19(7), 2006.
- [30] Wieringa R.J. Meyer J.-J. Ch., Dignum F.P.M. The paradoxes of deontic logic revisited : a computer science perspective. Technical Report UU-CS-1994-38, Utrecht University, Department of Computer Science, Utrecht, Netherlands, 1994.
- [31] E. Morin. *La Méthode : Tome 6, Ethique*. POINTS (Eds.), 2004.
- [32] E. Morin. *Introduction à la pensée complexe*. POINTS (Eds.), 2005.
- [33] A. Newell. The knowledge level. *Artificial Intelligence Journal*, 18 :87–127, 1982.
- [34] T. Powers. Deontological Machine Ethics. Technical report, American Association of Artificial Intelligence Fall Symposium 2005, Washington, D.C., 2005.
- [35] T. Regan and Singer P., editors. *Animal Rights and Human Obligations*. Prentice Hall, New-York, 1989.
- [36] Torrence S. Ethics and consciousness in artificial agents. *AI & Society*, 22 :495–521, 2008.
- [37] C. Tessier and Fr. Dehais. Authority management and conflict solving in human-machine systems. *AerospaceLab, The Onera Journal*, 4, 2012.
- [38] S. Tisseron. *Virtuel, mon amour. Penser, aimer, souffrir, à l'ère des nouvelles technologies*. Albin Michel, 2008.
- [39] S. Tisseron. *L'empathie au coeur du jeu social*. Albin Michel, 2010.
- [40] M. Tokoro. *Open Systems Science : From Understanding Principles to Solving Problems*. M. Tokoro (Ed.), IOS Press, 2010.
- [41] Alan Turing. Computing machinery and intelligence. *Mind*, 59 :433–460, 1950.
- [42] WJC. Verhagen, P. Bermell Garcia, P. Mariot, JP. Cotton, D. Ruiz, R. Redon, and R. Curran. Knowledge-based cost modelling of composite wing top cover (<http://peer.ccsd.cnrs.fr/peer-00649053>). *Int. Journal of Computer Integrated Manufacturing, IJCIM-0124 :TCIM–2010*, December 2010.
- [43] WJC. Verhagen and R. Curran. Ontological modelling of the aerospace composite manufacturing domain ([www.icas.org/icas\\_archive\\_cd1998-2010/icas2010/papers/483.pdf](http://www.icas.org/icas_archive_cd1998-2010/icas2010/papers/483.pdf)). In *Improving Complex Systems Today*, Advanced Concurrent Eng., pages 215–222. Springer, 2011.
- [44] F. Vexler, A. Berger, JP. Cotton, and A. Belloni. Éléments d'appréciation et d'analyse d'une base de connaissance : l'expérience industrielle d'Ardans. ([eric.univ-lyon2.fr/aide/actesAIDE\\_EGC2013ENLIGNE.pdf](http://eric.univ-lyon2.fr/aide/actesAIDE_EGC2013ENLIGNE.pdf)). *Actes Atelier AIDE @ EGC'2013, 13<sup>ème</sup> Conférence Francophone sur l'Extraction et la Gestion des Connaissances*, pages 59–72, Janvier 2013.
- [45] G. H. vonWright. Deontic logics. *Mind*, (60) :1–15, 1951.
- [46] Wendell Wallach and Colin Allen. *Moral Machines : Teaching Robots Right from Wrong*. Oxford University Press, USA, November 2008.
- [47] F. Zambonelli. Toward sociotechnical urban superorganisms. *IEEE Computer*, 45(8) :76–78, 2012.
- [48] T. Ågotnes, W. Hoek, and M. Wooldridge. Robust normative systems. In *AAMAS '08 Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, volume 2, pages 747–754, 2008.

# Compte-rendu des Journées Francophones sur les Systèmes Multi-Agents (JFSMA'12)

Honfleur, 17-19 octobre 2012

Rédigé par Bruno Mermet (Laboratoire Greyc-UMR 6072 et Université du Havre)

## 20èmes Journées Francophones des Systèmes Multi-Agents (JFSMA 12)

Depuis 1993, les Journées Francophones sur les Systèmes Multi-Agents (JFSMA) sont un moment annuel privilégié d'échanges scientifiques entre les chercheurs francophones dans le domaine des Systèmes Multi-Agents (SMA) et des chercheurs issus de différents domaines scientifiques (tels que l'intelligence artificielle, les sciences humaines, les systèmes distribués, la vie artificielle ou encore le génie logiciel).

La dix-neuvième édition des JFSMA s'est déroulée à Honfleur du 17 au 19 octobre 2012, après les éditions de Toulouse (1993), Grenoble (1994), Chambéry (1995), Port-Camargue (1996), Nice (1997), Nancy (1998), l'Île de la Réunion (1999), Saint-Etienne (2000), Montréal (2001), Lille (2002), Hammamet (2003), Paris (2004), Calais (2005), Annecy (2006), Carcassonne (2007), Brest (2008), Lyon (2009), Mahdia-Tunisie (2010) et Valenciennes (2011).

La participation était de 51 inscrits, stable par rapport à l'année dernière. Les JFSMA 2012 ont reçu le soutien des partenaires suivants :

- l'Université de Caen Basse-Normandie,
- le laboratoire GREYC - UMR 6072,
- la région Basse-Normandie,
- le département du Calvados,
- l'AFIA (Association Française pour l'Intelligence Artificielle),
- le GDR-I3.

Les Systèmes Multi-Agents sont un paradigme de premier choix pour l'étude et la conception de systèmes complexes au sein desquels plusieurs entités autonomes, les agents, contribuent à la réalisation d'une tâche selon un schéma d'organisation explicite ou non. Plutôt que d'être prescrit, le comportement global d'un SMA émerge du caractère adaptatif des agents et des capacités d'auto-organisation du système, tout en étant façonné par les contraintes spécifiques de la tâche à accomplir.

Les JFSMA permettent de faire le point sur les avancées dans les différents domaines de la recherche scientifique sur systèmes multi-agents. à l'occasion de la vingtième édition de cette conférence, nous avons choisi de revisiter des thèmes fondateurs de l'approche SMA à savoir les principes d'autonomie, d'ouverture et de co-évolution. En effet, à l'instar des sociétés humaines ou des systèmes biologiques, les SMA sont des systèmes ouverts, de part la nature des entités qui les composent et les échanges qu'elles entretiennent avec un environnement changeant. L'autonomie des agents et leurs interactions multiples sont l'essence même de la dynamique de ces sociétés au sein desquelles l'un s'enrichit grâce aux interactions qu'il entretient avec l'autre. La co-évolution des agents et de l'environnement assure la robustesse, l'adaptation aux changements, la pertinence de la solution et la viabilité du système. Ceci constitue l'originalité des fondements du paradigme des systèmes multi-agents.

Ces thèmes ont notamment été illustrés grâce à 2 conférenciers invités :

- Paul Richmond (Department of Automatic Control and Systems Engineering, University of Sheffield), nous a présenté comment utiliser efficacement des GPUs pour des simulations massives à base d'agents (Massive scale agent based simulation with FLAME GPU) ;
- François Bourdon (Laboratoire GREYC et IUT de Caen), nous a montré comment des agents mobiles pouvaient s'auto-organiser pour se répartir des ressources (Mobilité, coopération et auto-organisation dans les systèmes multi-agents).

L'ensemble des contributions présentées dans ces actes ne se réduit pas à ces seuls thèmes ; il reflète le dynamisme et la diversité des recherches actuelles sur les SMA dans la communauté francophone. Les actes de la présente édition comprennent les contributions sélectionnées par le comité de programme sur la base de leur évaluation par trois relecteurs. 18 articles ont été sélectionnés parmi les 36 soumissions. Lors de la conférence,

## COMPTE-RENDU DES JOURNÉES FRANCOPHONES SUR LES SYSTÈMES MULTI-AGENTS (JFSMA'12)

10 ont fait l'objet d'une présentation longue et 8 d'une présentation courte.

Le prix AFIA du meilleur papier a été décerné à Benoit Romito et François Bourdon pour : « Réduction de l'impact des fautes corrélées dans les réseaux pair-à-pair en utilisant des nuées d'agents mobiles », article particu-

lièrement bien rédigé et présenté également très clairement.

Les prochaines Journées Francophones sur les Systèmes Multi-Agents seront organisées sous la présidence de Salima Hassas (LIRIS, Lyon) à Lille, dans le cadre de la Plateforme Intelligence Artificielle, organisée par Philippe Mathieu (LIFL, Lille).

# Prix du meilleur article à JFSMA'12

Réduction de l'impact des fautes corrélées dans les réseaux pair-à-pair en utilisant des nuées d'agents mobiles

B. Romito [benoit.romito@unicaen.fr](mailto:benoit.romito@unicaen.fr) F. Bourdon [francois.bourdon@unicaen.fr](mailto:francois.bourdon@unicaen.fr)

Laboratoire GREYC, ENSICAEN, CNRS UMR 6072  
Université de Caen Basse-Normandie, France

## Introduction

Le stockage robuste de l'information dans les réseaux pair-à-pair repose sur la mise en place de schémas de l'information efficaces. En effet, la caractéristique principale de ce type d'architectures ouvertes est que les pairs peuvent se déconnecter à tout moment en privant le reste du réseau des données qu'ils hébergent. D'une manière générale, la tolérance aux fautes est obtenue par l'application de trois mécanismes :

1. introduire de la redondance dans les données (Weatherspoon et Kubiatowicz, 2002) pour tolérer les fautes ponctuelles. Cette redondance peut être obtenue par répllication ou par des codes d'effacement (Plank 1997) ;
2. assurer un placement adéquat de cette redondance pour en faciliter la supervision et la recherche (Giroire et al., 2009) ;
3. réparer la donnée lorsque trop de répliques ont été perdues (Dandoush, 2010). Les réparations sont indispensables pour assurer la durabilité du système.

De fait, chaque système définit son propre schéma de l'information et les performances de ce schéma en terme de tolérance aux fautes et de disponibilité sont évaluées sur des modèles de fautes. Cependant, la plupart des modèles de fautes employés comme c'est le cas dans CFS (Dabek et al., 2001) ou dans Farsite (Adya et al., 2002) font l'hypothèse que les fautes sont indépendantes. C'est-à-dire que les pairs se déconnectent indépendamment les uns des autres. Hors, en réalité, certains pairs partagent des traits communs les faisant covarier (Weatherspoon et al., 2002). Par exemple des pairs peuvent être situés dans le même bâtiment, ils peuvent partager les mêmes routeurs ou posséder les mêmes vulnérabilités. Héberger plusieurs répliques d'un même document sur des pairs corrélés expose le document à la perte simultanée de ces répliques alors que son modèle de fautes ne l'envisage pas. La disponibilité espérée s'en trouve alors faussée pouvant conduire dans le pire des cas à la perte de la donnée. Nous proposons dans cet article un placement des

répliques tenant compte des corrélations entre les pairs. Cette approche tire avantage de la mobilité d'un système multi-agents pour le stockage pair-à-pair que nous avons proposé dans (Pommier et Bourdon 2009) et dans (Pommier et al., 2010). Dans cette architecture, les documents sont fragmentés et transformés en nuées d'agents mobiles qui se déplacent dans le réseau. Après une étape de clustering qui regroupe les pairs fortement corrélés entre eux (Weatherspoon et al., 2002), nous proposons un algorithme de recuit simulé distribué (MinCor) qui optimise le placement des nuées sur ces clusters. Cet algorithme est en fait un filtrage des candidats dans le déplacement en flocking qui va guider les déplacements de chaque nuée afin qu'un maximum de ses fragments soient répartis sur des clusters distincts. Cette répartition a pour but de minimiser le nombre de pertes d'agents simultanées lorsque les fautes sont corrélées.

## Algorithme MinCor

Dans l'approche par clustering, les pairs sont regroupés en clusters. Chacun des clusters contient un ensemble de pairs qui sont corrélés. Nous faisons l'hypothèse, dans ces travaux, que les clusters sont disjoints et que la faute d'un pair a pour conséquence de déconnecter l'intégralité de son cluster (l'adéquation de ces hypothèses avec la réalité est discutée plus bas). Avec une telle structure, l'insertion des fragments d'une donnée doit être faite dans des clusters disjoints pour minimiser le nombre de pertes de fragments simultanées.

Alors que dans (Weatherspoon et al., 2002) la dissémination dans des clusters disjoints est réalisée au moyen d'un serveur central, notre approche consiste à décentraliser cette procédure, en tirant partie de la mobilité des documents évoluant en nuée. Cette recherche d'un placement de nuées sur des clusters disjoints est obtenue en contrôlant la direction du flocking à l'aide d'un algorithme de recuit-simulé distribué. Le recuit-simulé est une procédure itérative et non déterministe de recherche

des minima globaux d'une fonction. Cette fonction, que l'on appelle énergie par analogie avec la physique, donne une valeur de la qualité de la configuration dans laquelle le système se trouve actuellement. L'algorithme va chercher, par des changements de configurations successifs, à se trouver dans la configuration d'énergie minimale.

On donne la définition de l'énergie d'une nuée comme suit :

$$E(n) = \sum_{f \in n} \sum_{c \in C} c(f) + w \sum_{g \in G} |t(g, n)|^\gamma \quad (1)$$

L'énergie d'une nuée  $n$  dans une certaine configuration  $\alpha$  est une quantité qui est égale au coût d'hébergement de ses fragments (premier terme de la somme) ajoutée à la proportion de ses fragments sur des paires corrélés (second terme de la somme). Plus le nombre de fragments d'une même nuée sur le même cluster (le terme  $|t(g, n)|^\gamma$ ) est élevé et plus l'énergie résultante est élevée. La fonction  $t(g, n)$  retourne l'ensemble des paires du cluster  $g$  qui hébergent les agents de la nuée  $n$ . La constante  $\gamma \in R^+$  permet donc d'affecter une pénalité aux clusters contenant beaucoup de fragments de la même nuée.

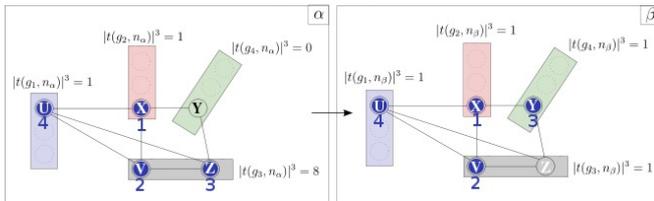


Figure 1 : illustrant un changement de configuration.

Comme le montre la Figure 1, le déplacement de l'agent 3 du pair Z au pair Y est souhaitable car elle diminue la valeur d'énergie du système et correspond à une meilleure répartition sur des clusters disjoints. La perturbation engendrée sur le système par ce changement de configuration est localisée dans les clusters source et cible du déplacement. Par conséquent, la fonction d'énergie est distribuée et peut être calculée localement par chaque agent. Cette propriété nous donne alors un schéma d'algorithme décentralisé, se basant sur le déplacement d'un seul agent à la fois, pour contrôler le flocking.

## Résultats

Nous avons conduit un ensemble d'expériences par la simulation afin d'évaluer cet algorithme. La première évaluation que nous proposons mesure la qualité du placement d'une nuée sur l'ensemble des clusters. Cette analyse consiste à comparer l'énergie d'une nuée se déplaçant avec l'algorithme MINCOR et celle d'une nuée se déplaçant avec l'algorithme de déplacement classique aléatoire.

Nous avons constaté que les nuées MINCOR trouvent un placement d'énergie minimum alors que les nuées classiques ont une énergie moyenne beaucoup plus élevée.

Le résultat de la première expérience atteste que l'algorithme trouve le minimum de  $E$ . Cependant, est-ce que cette fonction d'énergie permet bien de réduire l'impact des fautes corrélées sur le système. Nous évaluons, dans un second temps, le gain à utiliser cet algorithme en comparaison à un placement statique. Ce gain est mesuré en comptant le nombre moyen de pertes simultanées subies par chacune des nuées exécutant les différents algorithmes de placement. Il en ressort que les nuées MINCOR subissent moins de pertes simultanées de fragments que les autres placements et, par conséquent, sont plus tolérantes aux fautes corrélées.

## Conclusion et travaux futurs

Au delà du cadre applicatif des fautes corrélées, nous avons donné un algorithme permettant de diriger des nuées d'agents de manière asynchrone et décentralisée. Ce schéma d'algorithme reposant sur la mécanique du recuit-simulé n'est réalisable que si la fonction d'énergie est distribuée et calculable localement par chaque agent. Dans le cadre des fautes corrélées, nous avons montré que les nuées d'agents mobiles utilisant cet algorithme ont une meilleure tolérance aux fautes corrélées que les autres placements.

Dans des travaux futurs, notre réflexion se portera sur les hypothèses que nous avons posé. En effet, dans la réalité, un pair peut être corrélé avec plusieurs paires sur des sources de corrélations différentes. Par exemple, un pair  $A$  peut être corrélé avec un pair  $B$  parce qu'ils sont sur le même sous-réseau. Mais  $A$  peut également être corrélé avec un pair  $C$  parce qu'ils possèdent la même faille logicielle.  $B$  n'est alors pas nécessairement corrélé avec  $C$ . Dans ce cas, nous devons passer sur un modèle probabiliste des corrélations et pas sur un modèle dans lequel tous les paires d'un cluster sont déconnectés en même temps.

De même, il nous semble important d'analyser l'impact du coefficient  $w$  de la fonction d'énergie  $E$  pour conserver un bon placement sur les clusters tout en assurant un équilibre de la charge efficace.

## Bibliographie

Hakim Weatherspoon and John Kubiatowicz. 2002. Erasure Coding Vs. Replication : A Quantitative Com-

parison. In *Revised Papers from the First International Workshop on Peer-to-Peer Systems (IPTPS '01)*, 328-338.

James S. Plank. 1997. A tutorial on Reed-Solomon coding for fault-tolerance in RAID-like systems. *Software-Practice & Experience* 27, 9 (September 1997), 995-1012.

F. Giroire, J. Monteiro, and S. Pérennes. 2009. P2P Storage Systems : How Much Locality Can They Tolerate? In *Proceedings of the 34th IEEE Conference on Local Computer Networks (LCN)*, (October 2009), 320-323.

A. Dandoush. 2010. L'Analyse et l'Optimisation des Systèmes de Stockage de Données dans les Réseaux Pair-à-Pair. *Thèse de doctorat*, Université de Nice Sophia-Antipolis.

Frank Dabek, M. Frans Kaashoek, David Karger, Robert Morris, and Ion Stoica. 2001. Wide-area cooperative storage with CFS. *SIGOPS Operating Systems Review*. 35, 5 (October 2001), 202-215.

Atul Adya, William J. Bolosky, Miguel Castro, Gerald Cermak, Ronnie Chaiken, John R. Douceur, Jon Howell, Jacob R. Lorch, Marvin Theimer, and Roger P. Wattenhofer. 2002. Farsite : federated, available, and re-

liable storage for an incompletely trusted environment. *SIGOPS Operating Systems Review*. 36, SI (December 2002), 1-14.

Hakim Weatherspoon, Tal Moscovitz, and John Kubiatowicz. 2002. Introspective Failure Analysis : Avoiding Correlated Failures in Peer-to-Peer Systems. In *Proceedings of the 21st IEEE Symposium on Reliable Distributed Systems (SRDS '02)*, 362-.

Hugo Pommier and François Bourdon. 2009. Agents mobiles et réseaux pair-à-pair vers une gestion sécurisée de l'information répartie. *Revue d'Intelligence Artificielle*, 23(5-6), 697-718.

Hugo Pommier, Benoît Romito, and François Bourdon. 2010. Bio-inspired Data Placement in Peer-to-Peer Networks. Benefits of using Multi-agents Systems. In *Proceedings of the 6th International Conference on Web Information Systems and Technologies (WEBIST 2010)*, 319-324.

Benoît Romito, Hugo Pommier, and François Bourdon. 2011. Repairing Flocks in Peer-to-Peer Networks. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Volume 02 (WI-IAT '11)*, Pages 308-312.



## SOMMAIRE DU BULLETIN N° 79

<b>Editorial</b> .....	<b>3</b>
<b>Dossier Éthique et I.A.</b> .....	<b>4</b>
<b>Compte-rendu des Journées Francophones sur les Systèmes Multi-Agents (JF-SMA'12)</b> .....	<b>24</b>
<b>Prix du meilleur article à JFSMA'12</b> .....	<b>26</b>

## CALENDRIER DE PARUTION DU BULLETIN DE L'AFIA

	<i>Hiver</i>	<i>Été</i>
Réception des contributions:	<b>15 décembre</b>	Réception des contributions: <b>15 juin</b>
Sortie le	<b>31 janvier</b>	Sortie le <b>31 juillet</b>
	<i>Printemps</i>	<i>Automne</i>
Réception des contributions:	<b>15 mars</b>	Réception des contributions: <b>15 septembre</b>
Sortie le	<b>30 avril</b>	Sortie le <b>31 octobre</b>