

INTRODUCTION

Ces dernières années, éthique et intelligence artificielle ont souvent été mises en confrontation, soulevant questions, discussions, controverses. La métamorphose numérique de notre société, le développement des sciences et technologies de l'information et des communications et plus particulièrement de l'intelligence artificielle (IA) mettent l'éthique au cœur des questions liées à l'utilisation et la mise en œuvre de ces technologies.

Il est évident que toutes ces techniques pénètrent profondément dans notre vie quotidienne. Elles induisent de nombreuses et importantes modifications dans les rapports humains, dans de multiples pans de la société, tels que le travail, les modes de communications, l'économie. De nouvelles questions sont soulevées. Par exemple, les réflexions relatives à la protection de la vie privée doivent être repensées face à la généralisation de la collecte, de l'exploitation de masses de données individuelles s'appuyant sur des techniques issues de l'IA.

Ainsi se pose tout un ensemble de questions liées à la conservation de ces données et le droit à l'oubli. Dans le même ordre d'idée, face aux progrès des capacités d'autonomie, de décision et d'action des agents logiciels ou robotiques, les questions liées à la responsabilité des sujets humains (informaticiens, développeurs, vendeurs, etc.) et des dispositifs informatiques sont reformulés en des termes nouveaux et délicats. L'autonomie est devenue une caractéristique essentielle de nombreuses applications d'IA qui impactent notre quotidien (commerce électronique, jeux interactifs, intelligence ambiante, assistance aux personnes, robotique sociale ou collective, robotique civile, véhicule autonome, etc.). Dans un tel contexte, la question d'une régulation ou d'un contrôle éthique des agents logiciels ou robotiques ainsi que des systèmes constitués d'humains et d'agents autonomes se pose avec de plus en plus d'acuité.

Les rapports nationaux¹ ou internationaux² ainsi que différents projets de recherche³ mettent de plus en plus en évidence les questionnements, défis et directions de recherche à ce sujet. Ils montrent clairement que ces questions sont par nature transverses à de multiples disciplines. Ce numéro spécial croise ainsi les questionnements et les approches afin de mettre en lumière l'avancée des réflexions sur ces sujets.

Les deux premiers articles dressent un cadre de réflexion des enjeux, des concepts liés à l'éthique du numérique et plus particulièrement de l'intelligence artificielle en lien avec différentes applications de ces technologies.

1 <http://cerna-ethics-allistene.org> et <https://www.cnil.fr>

2 <https://ethicsinaction.ieee.org/>

3 <http://ethicaa.org>

Le premier article « Éthique et durabilité : des critères pour concevoir des interactions entre l'homme et les systèmes d'intelligence artificielle. Articuler les perspectives des législateurs à une vision contextuelle de l'éthique : le cas de l'expérience vécue des usagers de l'IA vocale » par D. Lahoual et M. Fréjus, propose une vision de l'éthique située et systémique en lien avec la notion de durabilité. Les auteurs proposent de privilégier la conception d'interactions éthiques et durables entre humains et systèmes fondés sur les techniques d'IA. Cette vision concilie contextes d'utilisation, points de vues des utilisateurs et développement de l'humain en s'appuyant sur des critères de conception conditionnant acceptabilité et appropriation d'un système d'IA. Les auteurs illustrent ces propositions à partir de retours d'expériences sur les usages de technologies vocales dans les sphères domestiques et professionnelles.

Dans une même démarche de clarification et d'analyse des technologies d'IA, M. Pégny et I. Ibnouhsein s'intéressent à l'usage du concept de « transparence » dans le débat public et scientifique portant sur les algorithmes. Leur article intitulé « Quelle transparence pour les algorithmes d'apprentissage machine ? » applique la taxonomie des usages de ce concept qu'ils proposent au cas particulier de l'apprentissage machine. Cette taxonomie met en évidence deux familles d'usages fondamentaux portant d'une part, sur des propriétés épistémiques des programmes telles qu'intelligibilité et explicabilité et, d'autre part, sur des propriétés normatives de leur usage telles que loyauté et équité.

Les deux articles suivants prennent un tout autre angle d'attaque de l'éthique et de l'IA en proposant des modèles formels visant à mettre en œuvre des comportements éthiques au sein de systèmes artificiels.

F. Berreby, G. Bourgne et J.-G. Ganascia se placent dans le contexte d'un agent raisonnant sur la décision en tenant compte de critères éthiques. Leur article, intitulé « Cadre déclaratif modulaire d'évaluation d'actions selon différents principes éthique », propose un cadre logique modulaire pour représenter et raisonner sur une variété de théories éthiques. Ce cadre présente quatre types de modèles interdépendants, implémentés en Answer Set Programming, qui permettent à l'agent d'évaluer son environnement, de raisonner sur sa responsabilité et de faire des choix éthiquement informés.

T. Vallée, G. Bonnet et T. de Swarte introduisent une dimension collective à l'éthique, dimension présente dans de nombreux domaines applicatifs où plusieurs agents artificiels doivent interagir, décider conjointement et coopérer. Leur article, intitulé « Modélisation de valeurs humaines : le cas des vertus dans les jeux hédoniques », s'intéresse au contexte particulier où un agent doit non seulement tenir compte de critères éthiques au regard de ses propres objectifs mais aussi sur la manière dont il coopère et tient compte des objectifs des autres agents. Les auteurs proposent un nouveau modèle de jeu hédonique de déviation pour la formation de coalition d'agents. Chaque agent décide ainsi des collectifs à former en s'appuyant sur une éthique fondée sur le respect de valeurs cardinales, appelée éthique des vertus.

Deux autres articles qui seront publiés dans le numéro suivant (5-6/2018) poursuivent cette investigation visant à introduire de l'éthique dans des systèmes artificiels. Ainsi L. Guizol et R. Baddoura dans l'article intitulé « Modèle d'éthique contextuelle pour agents artificiels » proposent un modèle pour représenter un problème éthique en termes de valeurs, vertus, règles, conséquences, intention ou acte à partir du contexte individuel, culturel, social et légal entrant dans la définition du contexte éthique d'un agent. R. Demolombe s'intéresse à l'analyse des interactions entre agents institutionnels, humains, logiciels et matériels. Cet article, intitulé « Modéliser les interactions entre agents : un prérequis pour analyser l'éthique des systèmes complexes » propose une formalisation en logique modale de ces différentes interactions en mettant l'accent sur les règles de raisonnement différentes pour capturer la multiplicité de ces interactions.

Nous remercions les auteurs et les membres du comité de relecture de ce numéro. Merci à tous ces collègues pour leur travail minutieux qui nous aura permis de réunir les articles que nous mettons en lumière dans ce numéro spécial consacré à éthique et intelligence artificielle.

GRÉGORY BONNET

GREYC, Université de Caen Normandie

EMILIANO LORINI

IRIT, CNRS

OLIVIER BOISSIER

Laboratoire Hubert Curien, École des Mines de Saint-Etienne

COMITÉ DE LECTURE DE CE NUMÉRO

Carole Adam – LIG, Université Joseph Fourier

Annie Blandin – IODE, IMT Atlantique

Gauvain Bourgne – LIP6, Sorbonne Université

Armelle Brun – LORIA, Université de Lorraine

Raja Chatila – ISIR, Sorbonne Université

Frédéric Cuppens – Télécom Bretagne, IMT Atlantique

Richard Dapoigny – LISTIC, Université de Savoie

Jean Paul Delahaye – LIFL, Université Lille 1

Robert Demolombe – IRIT

Laurence Devillers – LIMSI, Sorbonne Université

Amal El Fallah Seghrouchni – LIP6, Sorbonne Université

Catherine Faron-Zucker – I3S, Université de Nice Sophie Antipolis

Flora Fisher – CIGREF

Sébastien Gambs – IRISA, Université Rennes 1

Jean-Gabriel Ganascia – LIP6, Sorbonne Université

Christophe Garion – Onera

Olivier Georgeon – LIRIS, Université Catholique de Lyon

Alexei Grinbaum – LARSIM, CEA

Domitile Lourdeaux – Heudiasyc, Université de Technologie de Compiègne

Nicolas Maudet – LIP6, Sorbonne Université

Alexandre Monnin – Inria

Nathalie Nevejans – Centre Droit, Éthique & Procédures, Université d'Artois

Laurent Perrussel – IRIT, Université Toulouse 1

Guillaume Piolle – CentraleSupélec

Laurent Vercouter – LITIS, INSA de Rouen

Séréna Villata – CNRS, I3S

Catherine Tessier – Onera