Modélisation du respect de la vie privée dans des raisonnements éthiques formels

Pablo Rauzy <pr@up8.edu>
Université Paris 8 / LIASD

Gauvain Bourgne <gauvain.bourgne@lip6.fr>
Sorbonne Université / LIP6

Contexte

Le projet ReComp (*Realtime Compliance Mechanism for AI*) réuni des partenaires de trois pays : LIASD et LIP6 en France, IAI [1] en Allemagne, et NII [2] au Japon. Il a pour but d'améliorer la fiabilité des systèmes de décisions automatisés par la mise en œuvre de mécanismes de vérification en temps réel de la conformité aux normes éthiques et juridiques des solutions proposées par les IA.

L'éthique modulaire. À l'heure actuelle, les machines et les logiciels (agents) deviennent de plus en plus autonomes et agissent de plus en plus sans être contrôlés par des utilisateurs ou des opérateurs humains. C'est pourquoi, la question de doter ces agents autonomes de comportements éthiques se pose. En 2018, Berreby, Bourgne, et Ganascia [1] proposent un cadre logique nouveau et modulaire pour représenter et raisonner sur une variété de théories éthiques, sur la base d'une version modifiée du calcul des évènements, implémentée en Answer Set Programming. Le processus de prise de décision éthique est conçu comme une procédure en plusieurs étapes, capturée par quatre types de modèles interdépendants qui permettent à l'agent d'évaluer son environnement, de raisonner sur sa responsabilité, et de faire des choix éthiquement informés.

L'ambition est double : d'une part permettre la représentation systématique d'un nombre illimité de processus de raisonnements éthiques, à travers un cadre adaptable et extensible ; d'autre part éviter l'écueil trop courant d'intégrer directement l'information morale dans l'engin de raisonnement général sans l'expliciter, alimentant ainsi les agents avec des réponses atomiques qui ne représentent pas la dynamique sous-jacente.

L'objectif est de déplacer de manière globale le processus de raisonnement moral du programmeur vers le programme lui-même, en séparant en modules distincts d'une part ce qui concerne la prise de décision, l'éthique, et la mise en œuvre des actions, et d'autre part séparer clairement dans les "couches" du module éthique ce qui est général de ce qui est spécifique à un domaine, une application, voire un scénario.

La *privacy* par le contrôle. Le droit à la vie privée ne peut plus se résumer au "droit d'être laissé tranquille". Aujourd'hui il s'agit avant tout de laisser aux utilisateur·ices de systèmes d'informations le contrôle sur leurs données personnelles. Un cadre formel permettant la modélisation, la caractérisation, et l'évaluation de ce contrôle a été proposé en 2018 par Le Métayer et Rauzy : Capacity [3]. Trois axes de contrôle ont été identifiés dans la littérature juridique [2] :

— la capacité à effectuer des actions sur ses données personnelles,
— la capacité à empêcher des tiers d'effecteur des actions sur ses données personnelles, et
— la capacité à être informé des actions effectuées par des tiers sur ses données personnelles.

Dans Capacity, il y a des *agents* qui peuvent effectuer des *actions* (qui sont des *opérations* sur des *ressources* dans un *contexte* donné). Le contrôle est alors modélisé comme une *exigence*, qui exprime par une relation logique simple des contraintes sur des actions réalisables par des agents. Cette unique relation permet de modéliser les trois axes de contrôles cités ci-dessus.

La simplicité et la généricité de Capacity ne sont pas un frein à l'étude de systèmes tout à fait concrets. Capacity donne une sémantique aux exigences au travers de propriétés abstraites de traces. Ces propriétés abstraites peuvent être concrétisées pour un système donné après avoir identifié les agents, les ressources, les opérations, les contextes, et les évènements concrets qui composent les traces dans le système en question.

1. Institut Fur Angewandte Informatik
2. National Institute of Informatics

Objectifs

Des correspondances entre les modules du cadre éthique (actions concrète, éthique générale, règles spécifiques à une application ou un scénario) et les différents niveaux d'abstraction de Capacity (traces concrètes, propriété abstraites de traces, exigences) ont déjà été identifiées.

Le premier objectif est d'intégrer, en se basant sur ces correspondances, la notion de *privacy as control* dans une généralisation du cadre éthique. Cela permettra dans un second temps d'établir des spécifications éthiques de bonnes pratiques de conception vis-à-vis du traitement des données personnelles (*privacy by design*) et d'aider à la traduction dans un cadre formel de lois spécifiques comme le RGPD (spécifications juridiques). Les mécanismes de vérifications en temps réel développés au sein du projet ReComp pourront alors bénéficier de ces spécifications, permettant ainsi d'étendre son champ d'action à la protection de la vie privée.

Un autre axe de travail envisagé consiste en la construction d'une ontologie des données personnelles et des liens entre les différents types de données personnelles. L'objectif de ce travail est de faciliter le développement des spécifications concrètes des types de ressources dans le cadre formel de Capacity.

Profil du poste

Candidatures. Le ou la candidat·e idéal·e devra avoir des appétences pour la formalisation et l'abstraction, ainsi que des compétences en programmation, y compris en programmation logique (*Answer Set Programming*).

Seront également appréciés des connaissances sur les ontologies, ainsi que des intérêts personnels pour la philosophie, l'éthique, la politique, et la privacy.

Les candidatures sont à envoyer à Pablo Rauzy `<pr@up8.edu>` avec un CV et une lettre de motivation.

Accueil. Le poste est ouvert dans l'équipe PASTIS [3] du LIASD à l'Université Paris 8 (Saint-Denis, 93). Le travail se fera en collaboration étroite avec l'équipe ACASA [4] du LIP6 à Sorbonne Université (Paris).

Le poste est financé pour 12 mois (salaire net autour de 1950€), prévoit un équipement informatique, ainsi qu'un ou deux voyages (conférences ou séminaire du projet ReComp).

Références

[1] F. Berreby, G. Bourgne, and J.-G. Ganascia. Cadre déclaratif modulaire d'évaluation d'actions selon différents principes éthiques. *Revue des Sciences et Technologies de l'Information - Série RIA : Revue d'Intelligence Artificielle*, 2018.

[2] C. Lazaro and D. Le Métayer. Control over Personal Data : True Remedy or Fairy Tale ? *SCRIPTed*, 2015.

[3] D. Le Métayer and P. Rauzy. Capacity : an Abstract Model of Control over Personal Data. In *Conference on Data and Application Security and Privacy*, 2018.

3. `https://informatique.up8.edu/pastis/`
4. `http://www-poleia.lip6.fr/ACASA/`