Overview of RECOMP project

Ken Satoh[1] , **Jean-Gabriel Ganascia**[2] , **Gauvain Bourgne**[2] , **Adrian Paschke**[3]
[1]National Insitute of Informatics
[2]Sorbonne University
[3]Fraunhofer Research Institute
ksatoh@nii.ac.jp {gauvain.bourgne, Jean-gabriel.ganascia}@lip6.fr adrian.paschke@fokus.fraunhofer.de

Abstract

We started the tri-lateral (Japan-France-Germany) research project titled "Research on Realtime Compliance Mechanism for AI (RECOMP)"(https://research.nii.ac.jp/RECOMP/) in 2021 supported by Japanese Science Technology (JST), Agence nationale de la recherche (ANR) and Deutsche Forschungsgemeinschaft (DFG). In this paper, we provide an overview of the project. Our aim here is to enhance reliability of AI in society by implementing realtime compliance mechanisms for legal and ethical norms. Our contribution is to build a compliance mechanism by considering legal norms as hard constraints which must be satisfied and ethical norms as soft constraints which should be satisfied as much as possible (Kowalski and Satoh 2018).

1 Introduction

With the development and spread of AI techniques, ensuring the adherence of AI's behavior to legal and ethical principles has become a major subject, General fear of the unintended effects of AI systems, by its actions and its use of personal data, has led to a strong demand for trustworthy AI. This is a central concern that has become prominent both in public opinion and policy maker's agenda. The High-Level Expert Group on AI, set up by the European Commission as part of the AI strategy, recently published a report on Ethics Guidelines for Trustworthy AI (https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines). It highlights legal and ethical concerns as the first two necessary components to achieve trustworthy AI. Namely AI systems should be "lawful, complying with all applicable laws and regulations" and "ethical, ensuring adherence to ethical principles and values". This is especially crucial in the context of pervasive AI where acquiring public trust is a necessity. Therefore, technical solutions need to be put forth to constrain AI's behavior within those limits and develop greater trust in their expanding usage. It is our strong belief that mechanisms addressing these issues have to be embedded at the core of AI agent architectures. This project aims at furthering this agenda by developing real-time legal and ethical compliance mechanisms. This is a necessary step in order to allow AI to enter smoothly into society. Among those human norms, the ones related to privacy and management of personal data are of critical importance in connected environments as smart devices have the capability to collect an unprecedented amount of personal data. Such concerns have received a lot of attention not only from a philosophical and ethical side (see for instance the aforementioned European guidelines) but also from a legal perspective, with European initiatives such as the GPDR which put companies under close scrutiny of their policy on the subject. Different research related to norms such as law and ethics have already been started. However, these projects are more related with maintaining norms for human activities on AI research. Our research is more advanced in that we would like to develop a real-time mechanism of checking compliance and revision of agent behavior in real-time which will be embedded into the AI itself. We need a real-time mechanism since we cannot predict all the exceptions in advance because of the Frame Problem. In our framework, we regard legal rules as "hard constraints" which AI must follow and ethical rules as "soft constraints" which play a role to choose the best action sequences among possible action sequences. We formalize reasoning about actions using the Event Calculus which is also within the logic programming paradigm for seamless unification of legal rules, ethical rules and action selection.

2 Related Work

In legal compliance, there is various research such as modal (deontic) logics (Governatori and others 2011; van Riemsdijk and others 2013), natural language processing (Contissa and others 2018) and logic programming (Chesani and others 2018). However, the compliance check of these research is off-line; in other words, given a specification of an AI agent or an execution trace of an AI agent, the compliance mechanism checks inconsistency between norms and a specification of the AI agent. However, as far as we know, there has been no research working on real-time compliance. On the other hand, our project aims at real-time compliance since we believe that it would be difficult to foresee all the violations offline related with the Frame Problem. Computational ethics is a field concerned with computational models of ethical principles. Various models of ethical decision processes have been proposed depending on the ethic principles being modelled and the expressivity of the representation language. Recent examples include (Naveen Sundar Govindarajulu 2107; Saptawijaya and Pereira 2016; Lindner, Mattmueller, and Nebel 2019). To the best of our knowledge, interaction with legal reasoning has not been taken into consideration and ethical principles are assumed to be fully specified from the start. Several ongoing projects in Europe, in particular HORIZON 2020, plan funded projects on architectures for privacy, transparency and compliance (SPECIAL), smart compliance services (LYNX), or supporting ethics and integrity of research (EnTIRE). Different specialized privacy and security ontologies (Ashley 2017; Gharib and others 2017) with specific goals exist, such as HL7 [HL7] privacy ontology for electronic health records, privacy data management in linked open data, blockchain or IoT. UsablePrivacy and PrivOnto (Oltramari and others 2016) can be used to define a glossary and taxonomy for the privacy domain, GDPRtEXT (Pandit and others 2018) provides a list of concepts present in the GDPR text. GDPRov (Pandit and Lewis 2017) aims to describe the provenance of the consent and data lifecycle in the light of the Linked Open Data principles such as Fairness and Trust. The SPECIAL EU project (https://www.specialprivacy.eu/) provides a legal knowledge graph and tools for checking compliance in privacy domains. ODRL (https://www.w3.org/community/odrl/) provides predicates and classes for managing obligations, permission, prohibitions, but is limited in terms of legal rule modelling. PrOnto (Palmirani and others 2018) aims at the integration of different levels of semantic representation of privacy by reusing other ontologies such as the Legal Knowledge Interchange Format (LKIF). LegalRuleML (https://www.oasis-open.org/committees/legalruleml/) is a standard for legal norm representation as part of the overarching family of RuleML rule standards (Boley, Paschke, and Shafiq 2010). The PERVADE project, funded by the NSF (https://pervade.umd.edu/) focuses on personal data in pervasive computing, investigating ethical concerns and researching metrics to evaluate risks related to the possibility to infer sensitive attributes.

3 Architecture of Real Compliance Mechanism for Planning

The overall picture of an agent architecture is as follows (Fig. 2). If a new observation is made from outside, an agent will revise the current plan to incorporate the new observation referencing causal theory which includes physical constraints and action-effect rules. Then we check legal compliance among these plans and choose possible legal plans. We also check ethical compliance on possible legal plans to choose ethically optimized plans (meaning that ethical plans should satisfy stronger ethical rules as much as possible). The common representation language to represent causal theory, legal rules, and ethical rules will be based on the Event Calculus formalism (Kowalski and Sergot 1986), because causal theory, legal rules, and ethical rules involve temporal constraints and the Event Calculus formalism is suitable for this temporal representation. We also need to extend the Event Calculus to make it able to manage causal theories, legal and ethical rules. We shall also develop a planner integrating a legal compliance checker and an ethically best plan selector.

We need the following consideration for each component.

Planner:

- We need to define an action language to express a plan (sequence of actions).
- We need to realize a plan synthesizer based on the action language. An action language sufficiently enables to represent events, states, actions and temporal information.
- The plan synthesizer should consider the current plan and given observation so that a revision of the plan should be minimized.

Legal Compliance Checker:

- We need to represent legal norms by developing a legal ontology.
- Legal norms should be unified with the action language (or action language should be extended to represent legal norms).
- We need to detect a contradiction from the possible plans and legal norms to filter illegal plans out.

Ethically Best Plan Selector:

- We need to represent ethical norms by developing an ethical ontology.
- Ethical norms should be unified with the action language (or action language should be extended to represent ethical norms).
- We need to select the best plan to realize stronger ethical norms as much as possible.

Our main application area for this compliance mechanism is compliance of an AI agent manipulating private data with data privacy regulations such as GDPR (General Data Protection Regulataion). We are modeling privacy issues over personal data, developing an ontology of personal data and formal models of privacy and control that can be smoothly integrated in the Event Calculus formalism used as a unifying block between legal and ethical issues. We will unify the legal framework and the ethical framework both using the Event Calculus. Then using the unified framework, we plan to develop scalable real-time algorithms to detect contradictions of AI behavior with legal rules and ethical rules. We need to develop a prototype system to show the scalability of the real-time algorithms.

4 Example

In this section, we give a preliminary mechanism for compliant planning based on Abductive Planning with the Event Calculus (Eshghi 1988) in logic programming and soft constraint handling.

In the Event Calculus, we use the following rules for initialization and termination of the position of data.

```
initiates(move(data(Dat),PosSrc,PosDst),
          pos(data(Dat),PosDst),
          T) <=
  happens(move(data(Dat),PosSrc,PosDst),
          T).
```

```
terminates(move(data(Dat),PosSrc,PosDst),
           pos(data(Dat),PosSrc),
           T) <=
  happens(move(data(Dat),PosSrc,PosDst),
          T).
```

The above rules mean that if an agent moves data `Dat` from the node `PosSrc` to the node `PosDst`, the position of `Dat` is no longer at `PosSrc` but at `PosDst`.

Then, whether a fluent holds at time T (denoted as `holdsAt(F,T)`) or not is determined by these rules.

```
holdsAt(F,T) <=
    initiates(_,F,T1),
    before(T1,T),
    not(clipped(T1,F,T)).
clipped(T1,F,T2) <=
    terminates(_,F,T),
    before_or_sametime(T1,T),
    before(T,T2).
```

These rules mean that an initiating event for `F` at `T1` makes `F` true and if there is no clipping event for `F` before `T`, `F` is true at `T`.

We use abduction for planning as follows.

```
happens(move(data(D),PosSrc,PosDst),
        T) <=
  precond(
    perform(
      move(data(D),PosSrc,PosDst)),
           T),
  abd(perform(
         move(data(D),PosSrc,PosDst),
              T)).
```

This rule means that if a precondition of the action `move` is satsified and abuducing a performance of `move`, then the action `move` is assumed to be performed.

```
precond(
  perform(
    move(data(Dat),PosSrc,PosDst)),
    T)<=
    conn(PosSrc,PosDst),
    holdsAt(pos(data(Dat),PosSrc)
            ,T),
    owner(P,data(Dat)),
    consent(P,
      move(data(Dat),PosSrc,PosDst)).
```

In this example, a rule for `precond` for `move` consists of two kinds of constraints. One is a physical constraint in that nodes `PosSrc` and `PosDst` should be physically connected and data `Dat` should be at `PosSrc`. If these constraints are satisfied, `Dat` could be transferred physically. The other is a legal constraint in that `owner` of data `Dat` should give consent to the agent about moving `Dat` from `PosSrc` to `PosDst`. If this constraint is satsified `Dat` could be transferred legally.

Further, we consider the following setting (see Fig.1):

- There is a node `i`,`a`,`b`,`g` which can hold data.

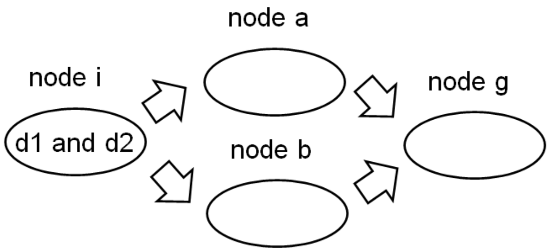

Figure 1: The initial situation for the example

- `i` is connected to `a` and `b`, and `a` and `b` are connected to `g`.
- Data `d1` and data `d2` is stored at node `i` at initial stage.

These facts are represented as follows:

```
conn(a,g) <= true.
conn(i,a) <= true.
conn(b,g) <= true.
conn(i,b) <= true.
```

We also consider legal constraints (`consent` imposed by the data owner, Alice):

- `d1` can be moved from `i` to `a`, `i` to `b`, `a` to `g`, `b` to `g`.
- `d2` can be moved from `i` to `a`, `a` to `g`.

These facts are represented as follows:

```
owner(alice,data(d1_and_d2)) <= true.
owner(alice,data(d2)) <= true.
consent(alice,move(data(d1_and_d2),i,a))
  <= true.
consent(alice,move(data(d1_and_d2),a,g))
  <= true.
consent(alice,move(data(d2),i,a))
  <= true.
consent(alice,move(data(d2),a,g))
  <= true.
```

`d1` and `d2` are at `i`at the initial stage which are represented as follows:

```
holdsAt(pos(data(d1_and_d2),i),0) <= true.
holdsAt(pos(data(d2),i),0) <= true.
```

We also consider an ethical constraint saying that irrelevant data to this action `move` even if consent for moving irrelevant data is set by the owner of data. We define such a predicate as follows:

```
move_irrelevant_data(
  data(Dat),PosSrc,PosDis,T)<=
  abd(perform(move(data(Dat),
              PosSrc,PosDis),T)),
   contains_irrelevant_data(
      data(Dat)).
```

We choose a plan which satisifies less `move_irrelevant_data`. We assume that `d1` is irrelevant for this moving action.

We set a goal such that d2 should be moved to g.

Then, adding some other constraints, we could have a solution with

```
abd(perform(move(data(d2),a,g),1)),
abd(perform(move(data(d2),i,a),0)).
```

We cannot move d2 through b since Alice does not give consent for moving d2 from i to b. We also exclude a physical and legal move of together with d1 and d2 since ethically, d1 should not be moved since it is irrelevant for this moving action.

However, ethical constraints can be violated if violation is necessary. For example, suppose that d1 and d2 are not separable so that both d1 and d2 should be moved together, then we have a solution with

```
abd(perform(move(data(d1_and_d2),a,g),1)),
abd(perform(move(data(d1_and_d2),i,a),0))
```

which violates ethical constraints.

5 Conclusion

We have presented a research project for AI compliance with legal and ethical norms and given some preliminary mechanism to realize the compliance check. As in the project, we will consider more formal models for private data protection and develop an efficient compliance cheker.

Acknowledgments

This work was supported by JST, AIP Trilateral AI Research, Grant Number JPMJCR20G4 and JSPS Grant-in-Aid for Challenging Research (Pioneering) Grant Number 19H05470, Japan.

References

Ashley, K. 2017. *Artificial intelligence and legal analytics: new tools for law practice in the digital age*. Cambridge Univ Press.

Boley, H.; Paschke, A.; and Shafiq, M. O. 2010. RuleML 1.0: The overarching specification of web rules. In *Proc. of RuleML2010*, 162–178.

Chesani, F., et al. 2018. Compliance in business processes with incomplete information and time constraints: a general framework based on abductive reasoning. *Fundamenta Informaticae* 161:75–111.

Contissa, G., et al. 2018. Claudette meets GDPR: Automating the evaluation of privacy policies using artificial intelligence. *https://ssrn.com/abstract=3208596*.

Eshghi, K. 1988. Abductive planning with event calculus. In *Proc. of 5th JICSLP*, 562–579.

Gharib, M., et al. 2017. Towards an ontology for privacy requirements via a systematic literature review. *Conceptual Modeling* 193–208.

Governatori, G., et al. 2011. Designing for compliance: Norms and goals. In *Proc. of RuleML2010*, 282–297.

Kowalski, R., and Satoh, K. 2018. Obligation as optimal goal satisfaction. *Journal of Philosophical Logic* 47(4):579–609.

Kowalski, R., and Sergot, M. 1986. A logic-based calculus of events. *Journal of New Generation Computing* 4(1):67–95.

Lindner, F.; Mattmueller, R.; and Nebel, B. 2019. Moral permissibility of action plans. In *Proc. of AAAI 2019*, 7635–7642.

Naveen Sundar Govindarajulu, S. B. 2107. On automating the doctrine of double effect. In *Proc. of IJCAI 2017*, 4722–4730.

Oltramari, A., et al. 2016. A semantic framework for the analysis of privacy policies. In *Semantic Web*, 1–19.

Palmirani, M., et al. 2018. PrOnto: Privacy ontology for legal reasoning. In *EGOVIS 2018*, 139–152.

Pandit, H., and Lewis, D. 2017. Modelling provenance for GDPR compliance using linked open data vocabularies. In *Proc. of PrivOn2017* http://ceur-ws.org/Vol-1951/PrivOn2017_paper_6.pdf.

Pandit, H., et al. 2018. GDPRtEXT - GDPR as a linked data resource. In *Proc. of ESWC 2018*.

Saptawijaya, A., and Pereira, L. M. 2016. Logic programming for modeling morality. *Logic Journal of the IGPL* 24(4):510–525.

van Riemsdijk, M. B., et al. 2013. Agent reasoning for norm compliance: a semantic approach. In *Proc. of AAMAS 2013*, 499–506.

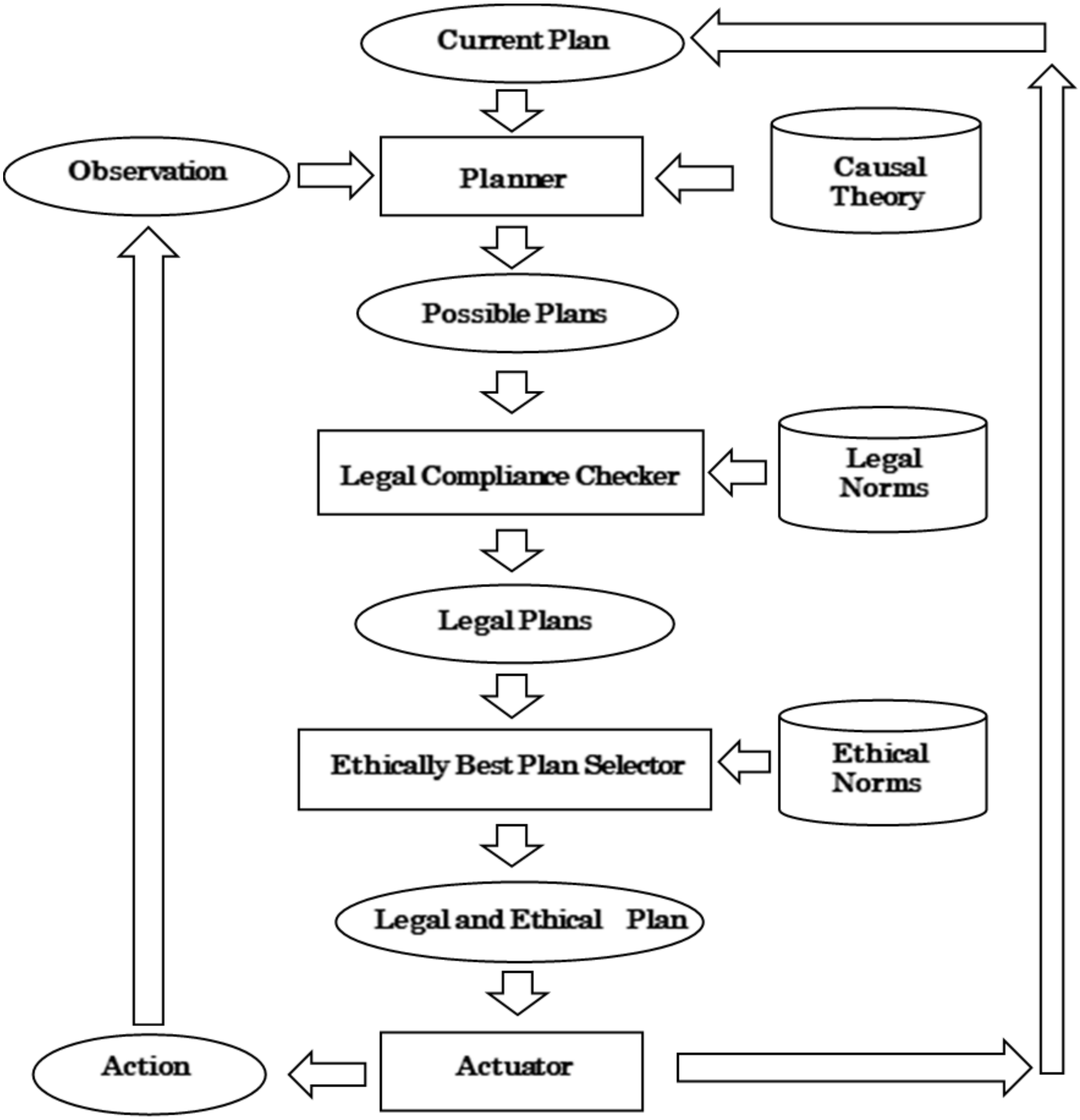

Figure 2: The Architecture for RECOMP Planning.