Realtime Compliance Mechanism for AI (RECOMP)

Adrian Paschke[1,2], Ken Satoh[3] and Jean-Gabriel Ganascia[4]

[1] Institut für Angewandte Informatik, Leipzig, Germany

[2] Fraunhofer FOKUS and Freie Universität Berlin, Berlin, Germany

[3] Center of Juris-Informatics, Research Organization of Information and Systems, Tokyo, Japan

[4] Sorbonne University, Paris, France

Abstract

The aim of the RECOMP project* is to enhance trustworthiness of norm-based AI by implementing a real-time compliance mechanism for legal and ethical norms.

1. Introduction

With the rapid evolution and spread of AI technologies in society, we can receive many benefits from AI. However, these same technologies also introduce new risks and negative consequences for individuals or society that threaten legal and ethical principles. Thus, we need to ensure that AI is compliant with these principles. This is a central concern that has become prominent both in public opinion and policy maker's agenda.

In the EU, there has been a proposal of "AI ACT" to ban AI systems that have an unacceptably high risk to create a clear threat to society, livelihoods, and rights of people and strongly regulate AI systems that have a high risk to be used in critical infrastructures and systems influencing human rights. Other AI systems are not regulated by this ACT but AI systems in general should be trustworthy, which means they should be lawful (respecting all applicable laws and regulations), ethical (adhering to ethical principles and values), and robust (both technically and considering their social environment). Therefore, technical solutions are needed to achieve this goal, and it is strongly believed that mechanisms addressing these issues should be embedded at the core of AI agent architectures.

2. The RECOMP Project

In this research we develop an overall architecture that combines legal and ethical Realtime Compliance (RECOMP*) mechanism to ensure the above conditions of trustworthy AI in the context of AI agent planning in multi-agent systems. [1,2]

As application example we consider the management of personal data. Concerns about privacy and protection of personal data 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 GDPR which put companies under close scrutiny of their policy on the subject. [3,4]

In our RECOMP 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 and we have a real-time compliance mechanism to check compliance even if there is a change of environment during the execution of a plan and we can modify the plan to be compliant with the new environment. [6,7,8]

3. RECOMP Concept

In this RECOMP research project, we develop the following modules. [1,5]
① Planner which can dynamically modify the current plan according to the change of environment (real-time replanning) and generate physically possible plans
② Legal compliance checker which filters illegal plans from output plans from the planner.
③ Ethically best plan selector which chooses ethically best plans among legal plans output from the legal compliance checker.
④ Execution module which performs the first action of one of the legal and ethically best plans.

We iterate this mechanism until the desired goal is obtained. (see the figure 1).
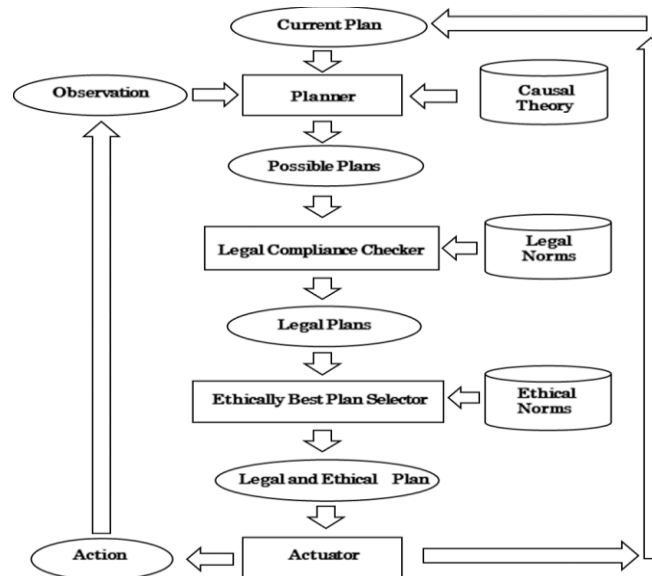
Figure 1: RECOMP Approach

For ①, we formalized causality which is a theoretical base of planning, and we implemented a planner which performs real-time replanning. For ②, we worked on formal representation of legal knowledge and implemented on legal compliance checker. For ③, we worked on formal representation of ethical knowledge and implemented ethically best plan selector in collaboration with Japanese team. For ④, we worked on multi-agent systems to combine planner, legal compliance checker and ethically best plan selector.

4. RECOMP Implementation

Planner:

As shown in the following figure, the planning agent module integrates a dynamic HTN planner, a legal checker, an ethical checker, and an action executor (Prova).
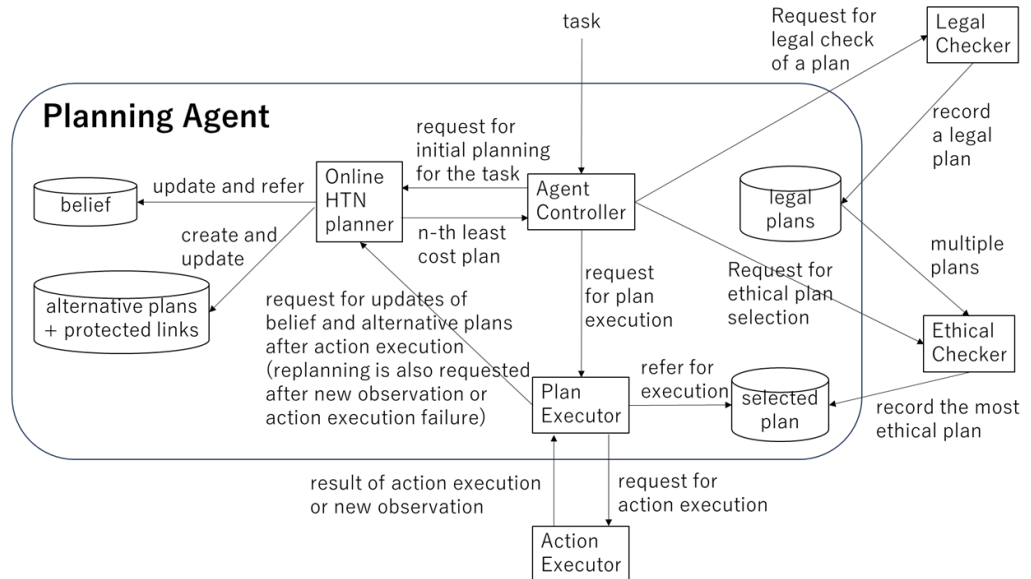
Legal Checker:

      The role of the legal checker is to check whether a given plan by the planning agent is legal. Because a plan is a list of actions, the legal checker checks all the actions and only decides that the plan is legal if all the actions are legal. The legal checker obtains actions from the plan received from the planning agent. The execution flow is shown in the following figure.
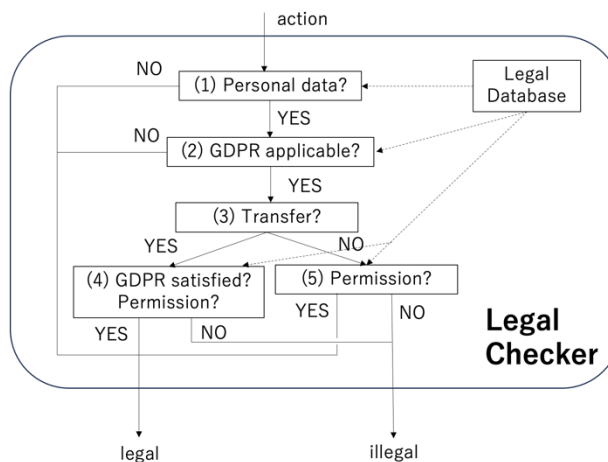
Ethically Best Plan Selector:

      The ethical check module takes as input a set of plans, then the plans are ordered according to several criteria and aggregated to obtain a final order. The first-order plan in the final aggregation is selected as the best. This mechanism is shown in the figure below. Certain properties are anticipated in the final order; for example, transitivity is essential to avoid counterintuitive results. That is why we use voting rules with desired properties that are well studied in social choice theory.
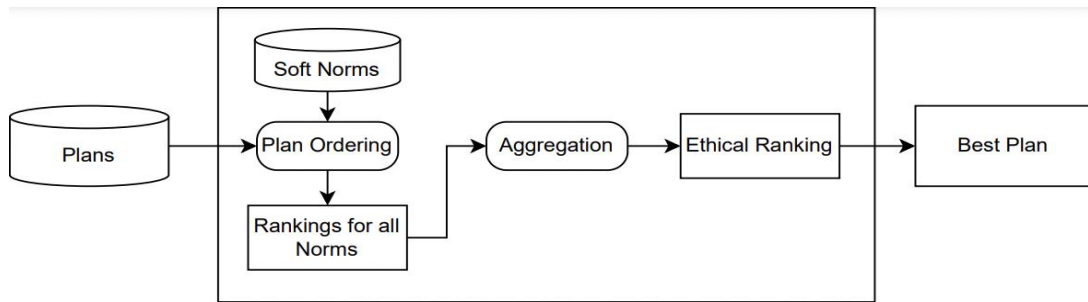
There are general properties that are desired in a voting rule, like Pareto's efficiency, monotonicity, etc. An important property in our case is the Condorcet principle, which means that a voting rule should select the alternative that beats every other alternative in pairwise comparisons as the winner. The Copeland's rule used in this method is an example of such rules; other rules such as max-min, etc. that satisfy the Condorcet principle can be used interchangeably. The lack of this property may cause a cyclic preference and lead to a loss of transitivity. The ethical checker uses domain knowledge to order plans and select the best one. It takes as input a list of legal plans, and orders them according to the defined soft norms. The orders can be seen as the votes of each norm on the input plans.

The Action Executor:

The purpose of the action executor consists of executing commands issued by the Plan Executor. The action executor is implemented in Prova, a (Semantic) Web rule language and a distributed (Semantic) Web rule engine that builds upon the ISO Prolog syntax and extends it with Java objects, typed variables, F-Logic-style slots, SPARQL and SQL queries. Prova supports reaction rule based workflows, event processing, and reactive agent programming. The communication between agents is realized with the message passing primitives that Prova provides. The communication between the Plan Executor and the Action Executor is performed by exchanging data in RuleML/XML interchange format. For the purpose of translating the commands (in Prolog form), a RuleML<->Prolog bidirectional translator server was developed, supporting the subset of the Prolog syntax necessary for the application. For instance, the following figure shows the Prova multi-agent execution implementation for data transfer according to a data privacy-preserving plan from the planning agent.
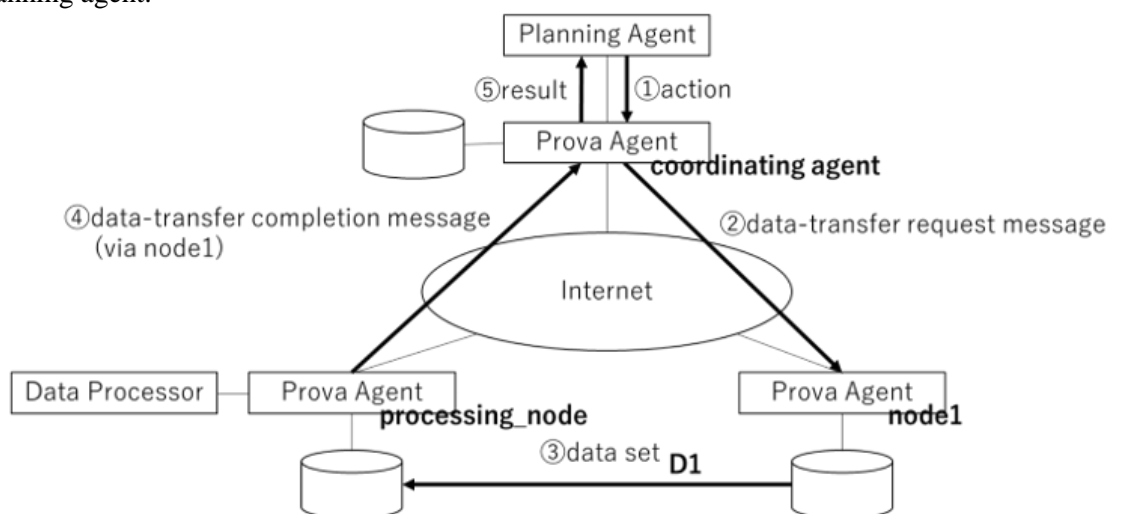
5. RECOMP Use Cases

The use case in [1] considers an international data transfer scenario, where an employment platform company provides job recommendation services for users. The company transmits personal data through its distributed servers to generate recommendations and deliver the result to users. On one hand, certain legal obligations from data protection regulations, in particular the European GDPR, apply to processing and transferring personal data. On the other hand, such use of personal data raises some ethical concerns about the privacy of users, the safety of personal data transfers, bias in the recommendations output, etc., that need to be respected. The company is using an automated process to handle data among its servers and wishes to ensure that legal obligations are met, and ethical criteria are respected as much as possible. The use case requires planning for data transfer and utilization considering legal and ethical norms. [9]

In [3] a GDPR-related use case for a distributed personal data wallet is described. A key GDPR concept is the concept of consent, meaning that the data controller, before initiating any data processing, is required to make an informed consent request to the user (data subject). Personal data wallet infrastructures are of particular interest in this context as their principle design goal is to provide a privacy and data security aware environment for exchanging personal data. The main concepts of the use case are data access from relaying parties or from users and providing consent that enables data sharing. Depending on consent providing or consent revoking actions, different possible states can emerge.

In [10] a use case for GDPR and DSA (Digital Services Act) compliance checking in moderated chat applications (here hate speech) is presented, utilizing the distributed personalized data wallets. The detection of hate speech or toxic content online is a complex and sensitive issue. While the identification itself is highly dependent on the context of the situation, sensitive personal attributes such as age, language, and nationality are rarely available due to privacy concerns. Additionally, platforms struggle with a wide range of local jurisdictions regarding online hate speech and the evaluation of content based on their internal ethical norms.

In [4] a use case for medical data access is presented. The use case focuses on medical data access, where actors such as a patient, doctors, researchers, and data controller act according to established legal rules. The implementation utilizes the concepts of consent and purpose, which are defined in the GDPR and similar legislations. It also includes emergency situations, where actors can gain access, overriding their default access rights.

The RECOMP implementation enables the legal and ethical compliance checking in such use cases in real-time. With the planning component the RECOMP framework also supports the dynamic replanning and selection of norm-compliant plans. This allows the executing agents to adapt to changes in the of environment during the execution of a plan.

References

[1] Hisashi Hayashi, Theodoros Mitsikas, Yousef Taheri, Kanae Tsushima, Ralph Schäfermeier, Gauvain Bourgne, Jean-Gabriel Ganascia, Adrian Paschke, and Ken Satoh. 2023. "Multi-Agent Online Planning Architecture for Real-Time Compliance." In *Proceedings of the 17th International Rule Challenge and 7th Doctoral Consortium @ RuleML+RR 2023 Co-Located with 19th Reasoning Web Summer School (RW 2023) and 15th DecisionCAMP 2023 as Part of Declarative AI 2023, Oslo, Norway, 18 - 20 September, 2023*, edited by Jan Vanthienen, Tomás Kliegr, Paul Fodor, Davide Lanti, Dörthe Arndt, Egor V. Kostylev, Theodoros Mitsikas, and Ahmet Soylu. Vol. 3485. CEUR Workshop Proceedings, Vol. 3485 (2023).

[2] Ken Satoh, Jean-Gabriel Ganascia, Gauvain Bourgne and Adrian Paschke: Overview of RECOMP project. *Proc. of International Workshop on Computational Machine Ethics (CME2021)* in conjunction with KR 2021, November 5 (2021).

[3] Ralph Schäfermeier, Theodoros Mitsikas, Adrian Paschke, Modeling a GDPR Compliant Data Wallet Application in Prova and AspectOWL. In: *Proceedings of the 16th International Rule Challenge and 6th Doctoral Consortium @ RuleML+RR 2022*, 26 - 28 September, 2022. CEUR Workshop proceedings, Vol-3229.

[4] Theodoros Mitsikas, Ralph Schäfermeier, and Adrian Paschke. 2023. "Modeling Medical Data Access with Prova." In: *Proceedings of the Seventeenth International Workshop on Juris-Informatics 2023 (JURISIN 2023) in association with the 15th JSAI International Symposia on AI (JSAI-isAI 2023), Kumamoto, Japan, 5-6 June, 2023,* edited by Ken Satoh and Katsumi Nitta.

[5] Theodoros Mitsikas, Ralph Schäfermeier, Yousef Taheri, Kanae Tsushima, Hisashi Hayashi, Jean-Gabriel Ganascia, Gauvain Bourgne, Ken Satoh, and Adrian Paschke. 2024. "Distributed Component Interoperation and Execution for Norm-Based Real-time Compliance" In *Proceedings of the 18th International Rule Challenge and 8th Doctoral Consortium @ RuleML+RR 2024, Bucharest, Romania, 16 - 18 September, 2024*. CEUR Workshop Proceedings, Vol. (2024).

[6] Gauvain Bourgne, Camilo Sarmiento, Jean-Gabriel Ganascia : "ACE modular framework for computational ethics : dealing with multiple actions, concurrency and omission", *International Workshop on Computational Machine Ethics*, Online event, France (2021).

[7] Yousef Taheri, Gauvain Bourgne and Jean-Gabriel Ganascia. "Modelling Integration of Responsible AI Values for Ethical Decision Making", *2nd International Workshop on Computational Machine Ethics*, KR2023 Workshop, Rhodes, Greece (2023).

[8] Camilo Sarmiento, Gauvain Bourgne, Katsumi Inoue and Jean-Gabriel Ganascia. "Action Languages Based Causality in Decision Making Contexts", in *proceedings of PRIMA (Principles and Practice of Multi-Agent Systems)*, pp 243-259 (2022).

[9] Hisashi Hayashi, Ken Satoh: Online HTN Planning for Data Transfer and Utilization Considering Legal and Ethical Norms: Case Study. ICAART (1) 2023: 154-164

[10] Jan Fillies, Theodoros Mitsikas, Ralph Schäfermeier, and Adrian Paschke. "A Hate Speech Moderated Chat Application: Use Case for GDPR and DSA Compliance", *International Workshop on AI Value Engineering and AI Compliance Meechanisms (VECOMP 2024)*, affiliated with the 27th European Conference on Artificial Intelligence (ECAI 2024) 19–24 October 2024, Santiago de Compostela, Spain.