Report from Dagstuhl Seminar 23151

Normative Reasoning for AI

Agata Ciabattoni[*1], John F. Horty[*2], Marija Slavkovik[*3], Leendert van der Torre[*4], and Aleks Knoks[†5]

1 **TU Wien, AT.** agata@logic.at
2 **University of Maryland – College Park, US.** horty@umd.edu
3 **University of Bergen, NO.** marija.slavkovik@uib.no
4 **University of Luxembourg, LU.** leon.vandertorre@uni.lu
5 **University of Luxembourg, LU.** aleks.knoks@uni.lu

Abstract

Normative reasoning is reasoning about normative matters – such as obligations, permissions, and the rights of individuals or groups. It is prevalent in both legal and ethical discourse, and it can – and arguably should – play a crucial role in the construction of autonomous agents. We often find it important to know whether specific norms apply in a given situation, and to understand why and when they apply, and why some other norms do not apply. In most cases, our reasons for wanting to know are purely practical – we want to make the correct decision – but they can also be more theoretical – as they are when we engage in theoretical ethics. Either way, the same questions are crucial for designing autonomous agents sensitive to legal, ethical, and social norms. This Dagstuhl Seminar brought together experts in computer science, logic (including deontic logic and argumentation), philosophy, ethics, and law with the aim of finding effective ways of formalizing norms and embedding normative reasoning in AI systems. We discussed new ways of using deontic logic and argumentation to provide explanations answering normative why questions, including such questions as "Why should I do A (rather than B)?", "Why should you do A (rather than I)?", "Why do you have the right to do A despite a certain fact or a certain norm?", and "Why does one normative system forbid me to do A, while another one allows it?". We also explored the use of formal methods in combination with sub-symbolic AI (or Machine Learning) with a view towards designing autonomous agents that can follow (legal, ethical, and social) norms.

Seminar April 10–14, 2023 – https://www.dagstuhl.de/23151
2012 ACM Subject Classification Computing methodologies → Artificial intelligence; Theory of computation → Logic; Computing methodologies → Multi-agent systems
Keywords and phrases deontic logic, autonomous agents, AI ethics, deontic explanations
Digital Object Identifier 10.4230/DagRep.13.4.1

[*] Editor / Organizer
[†] Editorial Assistant / Collector

Normative Reasoning for AI, *Dagstuhl Reports*, Vol. 13, Issue 4, pp. 1–23
Editors: Agata Ciabattoni, John F. Horty, Marija Slavkovik, Leendert van der Torre, and Aleks Knoks

Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Executive Summary

Agata Ciabattoni (TU Wien, AT)
John F. Horty (University of Maryland – College Park, US)
Marija Slavkovik (University of Bergen, NO)
Leendert van der Torre (University of Luxembourg, LU)

Normative reasoning – or, roughly, reasoning about such normative matters as obligations, permissions, and rights – is receiving increasing attention in several fields related to AI and computer science. There is an increase in its more traditional use in knowledge representation and reasoning, multiagent systems, and AI & law. However, it holds much promise and is also becoming more important in the context of the blooming fields of AI ethics and explainable AI. Accordingly, the interdisciplinary seminar *Normative Reasoning for Artificial Intelligence* brought together researchers working in knowledge representation and reasoning, multiagent systems, AI & law, AI ethics, and explainable AI to discuss ways in which normative reasoning can be used to make progress in the latter two disciplines.

While this Dagstuhl Seminar touched upon many different aspects of normative reasoning in AI, four topics received particular attention: (i) from AI & law to AI ethics, (ii) deontic explanations, (iii) defeasible deontic logic and formal argumentation, and (iv) from theory to tools.

From AI & law to AI ethics. AI & law is a field that is concerned with, on the one hand, laws that regulate the use and development of artificial intelligence and, on the other, the use of AI by lawyers and the impact of AI on the legal profession. In this field, normative systems are often used to represent and reason about the legal code. The seminar participants explored different ways in which ideas from AI & law can be used in the context of AI ethics.

Deontic explanations. This topic had to do with the use of formal methods, in general, and deontic logic and the theory of normative systems, in particular, to provide answers to why questions involving deontic expressions: "Why must I wear a face mask?", "Why is it forbidden for me to go out at night, although that other person is allowed to go out at night?", "Why has the law of privacy been changed in this way?". Deontic explanations have an essentially practical nature, which distinguishes them from (merely) scientific explanations. The concerns of scientific explanations focus on causality and uncertainty, whereas deontic explanations additionally include preferences, norms, sanctions, and actions. While causality and uncertainty are core concerns in explainable AI, in the context of our seminar, they played a relatively minor role. Instead, the seminar focused on the aspects of deontic explanations that are special to deontic explanations.

Defeasible deontic logic and formal argumentation. The third topic of the seminar had to do with the role of nonmonotonicity in deontic logic in general and the use of formal argumentation in particular. As is well known in the area of deontic logic, normative reasoning comes with its own set of benchmark examples and challenges, many of which are concerned with the handling of the so-called contrary-to-duty (CTD) reasoning and deontic conflicts. A whole plethora of formal methods have been developed to handle CTD and deontic conflicts, methods that go far beyond simple modal logics such as SDL (standard deontic logic). Furthermore, it is widely held that norms are defeasible and come with

exceptions and priorities. The seminar participants discussed the role of nonmonotonicity in deontic logic and the use of techniques from formal argumentation to define *defeasible* deontic logics.

From theory to tools. The fourth topic of the seminar concerned experimenting and implementing normative reasoning. One of the themes discussed had to do with integrating normative reasoning techniques with reinforcement learning (RL) in the design of ethical autonomous agents. Another theme that was discussed had to do with the automatization of deontic explanations. For example, in the recently introduced Logikey framework, it has been shown how Isabelle/HOL can be used as flexible interactive testbed for the design of domain-specific logical formalisms. Isabelle/HOL incorporates a number of automated tools that provide just-in-time feedback (counter-models, examples, proofs) to the formalization process. This feedback can be used to assess and reflect upon the theoretical properties of the system being designed/implemented. We can encode complex semantics in Isabelle/HOL as well as notions of argumentation (already partly done for abstract argumentation) so that Isabelle/HOL is turned into a reasoning system for those specific formalisms. What's more; notions of deontic explanations can be encoded and experimented with. Another key tool for automatize normative reasoning is analytic proof systems, which were also discussed in the seminar.

2 Table of Contents

3 Overview of Talks

3.1 Principles for a judgement editor based on Binary Decision Diagrams

Guillaume Aucher (University of Rennes, FR)

Joint work of Guillaume Aucher, Anthony Baire, Jean Berbinau, Annie Foret, Jean-Baptiste Lenhof, Marie-Laure Morin, Olivier Ridoux, François Schwarzentruber
Main reference Guillaume Aucher, Jean Berbinau, Marie-Laure Morin: "Principles for a Judgement Editor Based on Binary Decision Diagrams", Journal of Applied Logics -IfCoLog Journal of Logics and their Applications, Vol. 6(5), p. 33, 2019.
URL https://inria.hal.science/hal-02273483

We introduce the theoretical principles that underlie the design of a software tool which could be used by judges for making decisions about litigations and for writing judgements. The tool is based on Binary Decision Diagrams (BDD), which are graphical representations of truth-valued functions associated to propositional formulas. Given a type of litigation, the tool asks questions to the judge; each question is represented by a propositional atom. Their answers, true or false, allow to evaluate the truth value of the formula which encodes the overall recommendation of the software about the litigation. Our approach combines some sort of "theoretical" or "legal" reasoning dealing with the core of the litigation itself together with some sort of 'procedural' reasoning dealing with the protocol that has to be followed by the judge during the trial: some questions must necessarily be examined and sometimes in a specific order. That is why we consider extensions of BDD called Multi-BDD. They are BDD with multiple roots corresponding to the different specific issues that must necessarily be addressed by the judge during the trial. We illustrate our ideas on a case study dealing with French trade union elections which has been used throughout our project with the Cour de cassation. We also introduce the prototype developed during our project and a link with restricted access to try it out.

3.2 The moral disconnect in LLMs

Jan M. Broersen (Utrecht University, NL)

I will point out what is wrong with the moral behavior of LLMs like ChatGPT. Then I will ponder the question if we can actually solve the moral disconnect observed.

Large language model-based artificial conversational agents (like ChatGPT) can answer ethical questions. Just on the basis of that capacity, we may attribute a weak form of ethical knowledge to them. But do these models use this knowledge as a basis for their own ethical behaviour? I argue that cannot be the case. I will refer to this failure as the "ethical knowledge disconnect" of LLM-based agents. To understand the disconnect, we have to understand how ethical behavioural "guardrails" are implemented in systems like ChatGPT. I argue all methods currently employed do little to solve the disconnect. I wil also discuss how the disconnect may extend to non-ethical behaviours and should rather be seen as an instance of a more general knowledge disconnect. If that is the case, there are implications for making LLMs the basis for embodied agents. Finally, I will report on my attempt to expose the disconnect by trying to force ChatGPT into an ethical performative contradiction.

3.3 Machine ethics and precedent-based reasoning

Ilaria Canavotto (University of Maryland – College Park, US)

Joint work of Ilaria Canavotto, John Horty, Eric Pacuit
Main reference Ilaria Canavotto, John F. Horty: "Piecemeal Knowledge Acquisition for Computational Normative Reasoning", in Proc. of the AIES '22: AAAI/ACM Conference on AI, Ethics, and Society, Oxford, United Kingdom, May 19 – 21, 2021, pp. 171–180, ACM, 2022.
URL https://doi.org//10.1145/3514094.3534182

I will present research that I am carrying out in collaboration with John Horty and Eric Pacuit. We are exploring a hybrid approach to knowledge acquisition and representation for computational normative reasoning (a.k.a. machine ethics). Building on recent research in artificial intelligence and law, our approach is modeled on the familiar practice of decision-making under precedential constraint in the common law. I will first introduce a formal model of this practice (called the reason model of precedential constraint), showing how a body of normative information can be constructed in a way that is piecemeal, distributed, and responsive to particular circumstances. I will then discuss a possible application to the design of a robot childminder.

3.4 Data-driven norm revision

Mehdi Dastani (Utrecht University, NL)

Joint work of Davide Dell'Anna, Natasha Alechina, Fabiano Dalpiaz, Mehdi Dastani, Brian Logan
Main reference Davide Dell'Anna, Natasha Alechina, Fabiano Dalpiaz, Mehdi Dastani, Brian Logan: "Data-Driven Revision of Conditional Norms in Multi-Agent Systems", J. Artif. Intell. Res., Vol. 75, pp. 1549–1593, 2022.
URL https://doi.org//10.1613/jair.1.13683

Norm enforcement is a mechanism for steering the behavior of individual agents to achieve desired system-level objectives. Due to the dynamics of systems, however, it is hard to design norms that guarantee the achievement of the objectives in every operating context. In this work, we propose a data-driven approach to norm revision that synthesises revised norms with respect to a data set consisting of traces describing the behavior of the individual agents in the system. The proposed approach synthesises revised norms that are significantly more accurate than the original norms in distinguishing adequate and inadequate behaviors for the achievement of the system-level objectives.

3.5 Normative reasoning and the UK Highway Code

Louise A. Dennis (University of Manchester, GB)

Joint work of Joe Collenette, Louise A. Dennis, Michael Fisher
Main reference Joe Collenette, Louise A. Dennis, Michael Fisher: "Advising Autonomous Cars about the Rules of the Road", in Proc. of the Proceedings Fourth International Workshop on Formal Methods for Autonomous Systems (FMAS) and Fourth International Workshop on Automated and verifiable Software sYstem DEvelopment (ASYDE), FMAS/ASYDE@SEFM 2022, and Fourth International Workshop on Automated and verifiable Software sYstem DEvelopment (ASYDE)Berlin, Germany, 26th and 27th of September 2022, EPTCS, Vol. 371, pp. 62–76, 2022.
URL https://doi.org//10.4204/EPTCS.371.5

Our recent formalisation of the UK Highway Code has highlighted a number of ways normative reasoning interacts with it. Of particular interest are rules that implicitly defer to norms: e.g, "be considerate to other road users", many rules that have normative rather than legal force (the code distinguishes between rules that legally "must" be obeyed and rules that normatively "should" be obeyed), as well as rules which allow things which would be normatively impermissible (driving at night with the headlights off in well-lit areas). This opens up a defined area of computer reasoning in which the interaction of legal and normative rules can be studied.

3.6 Witnesses and explanations for answer set programming

Thomas Eiter (TU Wien, AT)

Joint work of Yisong Wang, Thomas Eiter, Yuanlin Zhang, Fangzhen Lin
Main reference Yisong Wang, Thomas Eiter, Yuanlin Zhang, Fangzhen Lin: "Witnesses for Answer Sets of Logic Programs", ACM Trans. Comput. Log., Vol. 24(2), pp. 15:1–15:46, 2023.
URL https://doi.org//10.1145/3568955

Answer Set Programming (ASP) is a popular declarative problem solving paradigm that has been widely applied in various domains. Given that answer sets are supposed to yield solutions to the original problem, the question of "why a set of atoms is an answer set" becomes important for both semantics understanding and program debugging. In this talk, we briefly consider recent work on answering such questions on disjunctive logic programs, as a basis for building explanations on top of ASP programs.

References

1 Yisong Wang, T. Eiter, Y. Zhang, and F. Lin. Witnesses for answer sets of logic programs. *ACM Transactions on Computational Logic*, 24(2), Apr. 2023. Article no. 15, 46 pp.

3.7 Machine learning with (logical) requirements

Eleonora Giunchiglia (TU Wien, AT)

Joint work of Eleonora Giunchiglia, Fergus Imrie, Mihaela van der Schaar, Thomas Lukasiewicz, Mihaela Stoian, Salman Khan, Fabio Cuzzolin

Machine learning models have revolutionised various fields by providing highly effective solutions to complex problems. However, their success comes at the cost of unexpected behaviours, which might violate known requirements expressing background knowledge about the problem at hand. This can have dramatic consequences, especially in safety critical scenarios (e.g., healthcare/autonomous driving). In this talk, I will first give an overview of the standard performance-driven machine learning development pipeline, and then I will present our proposed requirements-driven machine learning development process, highlighting its advantages. Finally, I will argue that it is desirable to use logic to express requirements, and I will briefly discuss how different neuro-symbolic methods have been developed to incorporate both norms (or soft constraints) and requirements (or hard constraints).

References

1 Eleonora Giunchiglia, Fergus Imrie, Mihaela van der Schaar, Thomas Lukasiewicz. *Machine Learning with Requirements: a Manifesto*. https://arxiv.org/pdf/2210.01597.pdf, 2023
2 Eleonora Giunchiglia, Mihaela Cătălina Stoian, Salman Khan, Fabio Cuzzolin, Thomas Lukasiewicz. *ROAD-R: The Autonomous Driving Dataset with Logical Requirements*. Machine Learning Journal, 2023
3 Eleonora Giunchiglia, Thomas Lukasiewicz. *Multi-Label Classification Neural Networks with Hard Logical Constraints*. Journal of Artificial Intelligence Research, 2021
4 Eleonora Giunchiglia, Mihaela Cătălina Stoian, Thomas Lukasiewicz. *Deep Learning with Logical Constraints*. IJCAI, 2022

3.8 Is the Chisholm paradox a paradox?

Guido Governatori (Tarragindi, AU)

Main reference Guido Governatori: "A Short Note on the Chisholm Paradox", in Proc. of the 4th International Workshop on MIning and REasoning with Legal texts co-located with the 32nd International Conference on Legal Knowledge and Information Systems (JURIX 2019), Madrid, Spain, December 11, 2019, CEUR Workshop Proceedings, Vol. 2632, CEUR-WS.org, 2019.
URL https://ceur-ws.org/Vol-2632/MIREL-19_paper_4.pdf

We advance an alternative version of the Chisholm Paradox and we argue that the alternative version (while logically equivalent to the original version), in its manifestation in the natural language, is not intuitively consistent. The alternative version of the paradox suggests some requirements for deontic logics designed for legal reasoning.

References

1 José Carmo and Andrew J. I. Jones. *Deontic Logic and Contrary-to-Duties*, Handbook of Philosophycal Logic (2nd edition), Volume , pages 265–343. Springer Netherlands, Dordrecht, 2002.
2 Rodrick M. Chisholm. Contrary-to-Duty Imperatives and Deontic Logic. *Analysis*, 24(2):33–36, 1963.

3 Guido Governatori. A short note on the Chisholm Paradox. Proceedings of the 4th International Workshop on MIning and REasoning with Legal texts, CEUR-Workshop Proceedings 2632.
4 John Horty. Deontic modals: Why abandon the classical semantics. *Pacific Philosophical Quarterly*, 95:424–460, 2014.
5 James E. Tomberlin. Contrary-to-duty imperatives and conditional obligation. *Noûs*, 15(3):357–375, 1981.
6 Lennart åqvist. Good Samaritan, contrary-to-duty imperatives, and epistemic obligations. *Noûs*, 1(4):361–379, 1967.

3.9 n problems for deontic logic for normative reasoning.

Guido Governatori (Tarragindi, AU)

Main reference Guido Governatori: "Thou shalt is not you will", in Proc. of the 15th International Conference on Artificial Intelligence and Law, ICAIL 2015, San Diego, CA, USA, June 8-12, 2015, pp. 63–68, ACM, 2015.
URL https://doi.org//10.1145/2746090.2746105

The original intent of this talk was to highlight some problems/issues a deontic logic has to address to capture normative (legal) reasoning (including some that might be controversial). However, after some of the previous presentation, the focus shifted to one of the issues, more specifically whether it is possible to use other logic, in particular Temporal Logic to model legal reasoning. I show that using Temporal Logic, more precisely, Linear Temporal Logic, as done by some work in the area of business process compliance, leads to some paradoxical results: either it is not possible to model some deontic aspects, or the outcome of the modelling contradicts expected legal outcome.

References

1 Guido Governatori. Thou Shalt is not You Will. Proceedings International Conference on Artificial Intelligence and Law 2015, pp. 63-68. doi: 10.1145/2746090.2746105
2 Guido Governatori and Mustafa Hashmi. No Time for Compliance. Proceedings EDOC 2015: pp. 9-18 doi: 10.1109/EDOC.2015.12

3.10 Deontic explanation: questions, dilemma's and choice

Joris Hulstijn (University of Luxembourg, LU) and Leendert van der Torre (University of Luxembourg, LU)

When a computer system takes decisions that affect people, they may demand an explanation. When the application involves norms, we need a deontic explanation. An deontic explanation is analyzed here as an answer to a why-question, relative to a normative system. Just like a who-question asks for persons, a why-question asks for reasons. We analyze differences and similarities of three kinds of semantics, that are formulated in terms of a partition of the set of possible worlds: (1) questions and answers, (2) moral dilemmas, and (3) see-to-it-that choice structures. The analysis is built on an analogy between providing an answer to a

question, resolving a moral dilemma and choosing an action. The role of the context in these types of semantics can be naturally analysed in a form of update semantics. In future work we hope to find constructions in the object language, to construct or reframe a question, a choice for action, or a dilemma, and ways of answering or resolving them.

3.11 The logic of second-order reasons

Aleks Knoks (University of Luxembourg, LU)

A normative reason is a consideration that counts either in favor of or against an action or attitude. A second-order normative reason, then, is a consideration that counts in favor of or against taking another consideration to be a normative reason. While some authors have questioned the existence of such reasons, others assign them very important roles. Thus, exclusionary or negative second-order reasons – that is, reasons against taking other considerations to be reasons – play a crucial role in Joseph Raz's account of practical reasoning. The primary goal of this talk is to show how second-order reasons and their normative effects can be captured in default logic. Starting with Horty s default logic-based model of the way reasons interact to support ought statements, I explain why one can't rest content with Horty's formalization of exclusionary reasons. Most importantly, it assimilates defeat by exclusionary reasons to canceling (or undercutting), doesn't do justice to the idea that excluded first-order reasons remain valid, and doesn't account for a distinct sense of "ought" grounded in first-order reasons. I discuss an alternative model, present an account of positive-second order reasons, and explore the model's predictions regarding the structure of even higher-order reasons and conflicts between them.

3.12 Moral planning agents

Emiliano Lorini (CNRS – Toulouse, FR)

The talk shows how non-classical logics with special emphasis on modal logic, epistemic logic and conditional logic can be used to represent and compare a rich variety of explanations of classifier systems; these include abductive, contrastive, counterfactual, objective vs subjective, and interactive explanations. The first part of the presentation will be devoted to explaining "white box" classifiers that are assumed to be perfectly known, while the second part will focus on "black box" classifiers about which the external observer has only partial knowledge. I will present proof-theoretic and complexity results for the involved logics and illustrate their expressiveness through concrete examples.

References

1 Liu, X., Lorini, E. (2023). A Unified Logical Framework for Explanations in Classifier Systems. Journal of Logic and Computation, 33(2), pp. 485-515
2 Liu, X., Lorini, E. (2022). A Logic of "Black Box" Classifier Systems. In Proceedings of the 28th Workshop on Logic, Language, Information and Computation (WOLLIC 2022), LNCS, volume 13468, Springer-Verlag, pp. 158–174

3 Liu, X., Lorini, E. (2021). A Logic for Binary Classifiers and Their Explanation. In Proceedings of the 4th International Conference on Logic and Argumentation (CLAR 2021), LNCS, volume 13040, Springer-Verlag, pp. 302-321.
4 Aguilera-Ventura, C., Herzig, A., Liu, X., Lorini, E. (2023). Counterfactual Reaasoning via Grounded Distance. In Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning (KR 2023), forthcoming.

3.13 How to implement cognitive and social properties of norms in robots: The promise of behavior trees

Bertram F. Malle (Brown University – Providence, US)

Joint work of Bertram F. Malle, Eric Rosen, Vivienne B. Chi, Dev Ramesh
Main reference Bertram F. Malle, Eric Rosen, Vivienne B. Chi, Dev Ramesh: "What properties of norms can we implement in robots¿' Proceedings of the 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023), August 2023.

Norms are indispensable for human communities, and so they will be for robot-human communities. We analyze some of the requirements for a robot to have norms and conform its actions to them. These requirements include both cognitive and social properties that human norms have. We examine which of these properties can be implemented in a robot's architecture and review some previous computational approaches. We then introduce a new one using behavior trees, argue for its promise to implement properties of norms, and discuss unsolved challenges.

3.14 Towards a mechanisation of the proof theory of normative reasoning

Xavier Parent (TU Wien, AT)

Joint work of Xavier Parent, Agata Ciabattoni, Nicola Olivetti
Main reference Agata Ciabattoni, Nicola Olivetti, Xavier Parent: "Dyadic Obligations: Proofs and Countermodels via Hypersequents", in Proc. of the PRIMA 2022: Principles and Practice of Multi-Agent Systems – 24th International Conference, Valencia, Spain, November 16-18, 2022, Proceedings, Lecture Notes in Computer Science, Vol. 13753, pp. 54–71, Springer, 2022.
URL https://doi.org//10.1007/978-3-031-21203-1_4

This work lays the groundwork for a (proper) mechanisation of normative reasoning, via the use of a so-called analytic sequent calculi. They are particularly useful for backward reasoning and deontic explanations. To answer a question of the form "why should I do X?", needed is to retrieve the path (the "proof") leading to this conclusion. Analytic calculi allow precisely this.

This is part of a bigger project aiming at developing analytic proof systems for deontic logic formalisms.

We consider a SoA formalism, the preference-based system E for conditional obligation due to Aqvist. Its key strength lies in its ability to resolve the CTD paradox. We provide an analytic calculus for it, the first of its kind. We also provide a terminating countermodel generation procedure in case of failure of proof search, and a complexity result (co-NP).

3.15 Deontic to description logics

Bijan Parsia (University of Manchester, GB)

Joint work of Bijan Parsia, E'leanor Turner, Ui Sattler

Deontic logic is a family of often propostional modal logics intended to capture certain kinds of moral reasoning, centrally those involving obligation. Deonitc logic has receive a great deal of attention from philosophical logicians and there has been some interest in developing implementations of reasoning procedures for various deontic logics. However, much of the work has rested on axiomatic approaches with the inference rules of necessitation and *modus pones*. While familiar and convenient for some communities, these not a standard basis for robust implementations.

Description logics are a widespread family of decidable logics, the core of which can be seen as notational variants of propositional modal logics. The description logic $\mathcal{SHROIQ}$ forms the logical foundation of the standardised ontology language OWL 2 DL. OWL 2 has broad and deep infrastructure including production quality reasoners, IDEs, services, repostiories, and so on.

Give the very expressivity and wide range of available tools, description logics are an attractive foundation for a translational approach to implementing reasoning services for deontic logics.

We present various translations of several logics of obligation and agency into OWL 2. While all the translations preserve the meaning (at least in the sense of always being able to recover all entailments) of the original, they have different characteristics which affect the performance of automated reasoning tasks and the utility of the results for users.

We also explore the benefits of an ontology-oriented approach to modeling deontic problems.

3.16 KI Wissen – Development of methods for integrating knowledge into machine learning in autonomous driving

Adrian Paschke (FU Berlin, DE)

Main reference Julian Wörmann, Daniel Bogdoll, Etienne Bührle, Han Chen, Evaristus Fuh Chuo, Kostadin Cvejoski, Ludger van Elst, Tobias Gleißner, Philip Gottschall, Stefan Griesche, Christian Hellert, Christian Hesels, Sebastian Houben, Tim Joseph, Niklas Keil, Johann Kelsch, Hendrik Königshof, Erwin Kraft, Leonie Kreuser, Kevin Krone, Tobias Latka, Denny Mattern, Stefan Matthes, Mohsin Munir, Moritz Nekolla, Adrian Paschke, Maximilian Alexander Pintz, Tianming Qiu, Faraz Qureishi, Syed Tahseen Raza Rizvi, Jörg Reichardt, Laura von Rüden, Stefan Rudolph, Alexander Sagel, Gerhard Schunk, Hao Shen, Hendrik Stapelbroek, Vera Stehr, Gurucharan Srinivas, Anh Tuan Tran, Abhishek Vivekanandan, Ya Wang, Florian Wasserrab, Tino Werner, Christian Wirth, Stefan Zwicklbauer: "Knowledge Augmented Machine Learning with Applications in Autonomous Driving: A Survey", CoRR, Vol. abs/2205.04712, 2022.
URL https://doi.org//10.48550/arXiv.2205.04712

AI-based processes are paving the way to fully automated autonomous driving. Up until now, the development of AI solutions has been purely driven by data. This data driven approach requires enormous amounts of data for the training and validation of AI functions, with the collection and processing of this data being very resource-intensive and expensive. In addition to the dependence on extensive amounts of data, data-based AI processes have

another weakness: they are still generally black-box models for which the decision making process cannot be directly reconstructed. In the talk I will report about the project "KI Wissen" (`https://www.kiwissen.de/`) and neuro-symbolic methods for integrating existing knowledge into the data-driven AI functions of autonomous vehicles (AVs) and vice versa for extracting interpretable symbolic knowledge from deep neural network models. In this talk I will specifically present an approach for extracting interpretable hierarchical rules from the learned AVs' deep neural networks and a neuro-symbolic architecture for a hybrid integration of deep neural networks, modelling the AVs' behaviour and situation information, with symbolic knowledge models for representing ontological domain and world knowledge for situation interpretation and for representing rule-based legal knowledge and norms for legal reasoning and compliance checks. The goal of the KI Wissen project is to create a comprehensive ecosystem for the integration of knowledge into the training and safeguarding of AI functions. By combining conventional data-based AI methods with the knowledge- or rule-based methods developed in the project, the basis for training and validating of AI functions will be completely redefined: This basis now includes not only data, but information, i.e., data and knowledge. The development from data- to information-based AI carried out in the project addresses the central challenges towards autonomous driving: the generalization of AI to phenomena with small data bases, the increase of the stability of the trained AI to disturbances in the data, the data efficiency, the plausibility check and the validation of AI-supported functions as well as the increase of the functional quality.

3.17 Legal explanations

Antonino Rotolo (University of Bologna, IT)

One fundamental question lies behind the distinction between justification and explanation in normative reasoning. In fact, we must notice that the tradition of legal logic and legal theory, in modeling legal decision-making, very often elaborate on various types of justification and takes this last concept as central, somehow maintaining that the idea of explanation depends on justification: while the explanation of a legal decision does not necessarily correspond to a justificatory reason for it, the opposite usually holds. The talk will offer some formal insights about this topic.

3.18 Combining Deep NLP with symbolic reasoning in automatic legal judgement

Ken Satoh (National Institute of Informatics – Tokyo, JP)

Joint work of Ha-Thanh Nguyen, Wachara Fungwacharakorn, Fumihito Nishino, Ken Satoh, Magumi Fujita
Main reference Ha-Thanh Nguyen, Wachara Fungwacharakorn, Fumihito Nishino, Ken Satoh: "A Multi-Step Approach in Translating Natural Language into Logical Formula", in Proc. of the Legal Knowledge and Information Systems – JURIX 2022: The Thirty-fifth Annual Conference, Saarbrücken, Germany, 14-16 December 2022, Frontiers in Artificial Intelligence and Applications, Vol. 362, pp. 103–112, IOS Press, 2022.
URL https://doi.org//10.3233/FAIA220453

We show how to combine deep NLP with nonmonotic reasoning in legal domain. We extract legally relevant facts from a case description written in natural language using deep NLP and input these facts into manually encoded articles in our legal logic programming language PROLEG to make legal judgement and produce explanation of the judgement.

3.19 What are social norms?

Kai Spiekermann (London School of Economics, GB)

Main reference Kai Spiekermann: "Review: Explaining Norms, Geoffrey Brennan, Lina Eriksson, Robert E. Goodin and Nicholas Southwood", Oxford University Press, 2013, Economics and Philosophy, vol. 31, no. 1.
URL https://www.kaispiekermann.net/blog-native/2015/6/9/book-review-explaining-norms-geoffrey-brennan-lina-eriksson-robert-e-goodin-and-nicholas-southwood

What are social norms? There is a surprising level of disagreement about this question in the literature. I compare Bicchieri's (2003) game-theoretic account with Brennan, Erikson, Goodin, and Southwood's account of norms as cluster of attitudes. Both accounts agree that real or perceived social practices and normative expectations play a central role. However, examples show that the different accounts can come apart.

4 Working groups

4.1 Explanation in case-based reasoning

Ilaria Canavotto (University of Maryland – College Park, US), John F. Horty (University of Maryland – College Park, US), Bijan Parsia (University of Manchester, GB), and Henry Prakken (Utrecht University, NL)

Computational models of legal precedent-based reasoning developed in the field of Artificial Intelligence and Law have recently been applied to the development of explainable AI methods. The key idea behind this approach is to interpret training data as a set of precedent cases; a model of precedent-based reasoning can then be used to build either an interpretable system for binary classification [1, 2] or an algorithm that generates post-hoc justifications for the

decisions of a machine learning system for binary classification [3]. This breakout session has been devoted to discuss a number of technical and conceptual questions concerning the framework for post-hoc justification proposed in [3].

References

1 Cocarascu, O. Čyras, K. and Toni, F. *Explanatory predictions with artificial neural networks and argumentation*. IJCAI/ECAI-2018 Workshop on Explainable Artificial Intelligence, pp. 26-32.

2 Čyras, K., Satoh, K. and Toni, F. *Explanation for case-based reasoning via abstract argumentation*. COMMA 2016, pp. 26–32.

3 Prakken, H. and Ratsma, R. *A top-level model of case-based argumentation for explanation: Formalisation and experiments*, Argument & Computation 13, 2022, pp. 159–194.

4.2 Justification and explanation

Ilaria Canavotto (University of Maryland – College Park, US), Pedro Cabalar (University of Coruña, ES), Thomas Eiter (TU Wien, AT), Joris Hulstijn (University of Luxembourg, LU), Aleks Knoks (University of Luxembourg, LU), Eric Pacuit (University of Maryland – College Park, US), Bijan Parsia (University of Manchester, GB), Henry Prakken (Utrecht University, NL), and Antonino Rotolo (University of Bologna, IT)

The problem of developing explainable AI methods is becoming increasingly central in AI. At the same time, the notions of explanation, explainability, and justification have been extensively investigated in philosophy, law, and social science. As a result, an increasing number of scholars from these disciplines is considering how to apply research in these fields to explainable AI. One problem of this interdisciplinary effort is that, more often than not, researchers from different fields have different understandings of what the task underlying explainable AI is (or should be) or what exactly "explanation" or "justification" mean when applied to AI systems. This working group aimed at identifying some key distinctions that could be used to build a unified conceptual framework for explainable AI. The discussion was split into two parts:

Part 1: Initial definitions. Most participants agreed that, when discussing explainable AI methods, it is helpful to introduce an initial distinction between explanation, explication, and justification. Although there was substantial disagreement about the exact definition of each notion, we agreed on the following prelimary characterization:

Explanation is about how a system reached a particular decision given a certain input and what the reason (motive, or cause) of the decision was. Explanation is important for bias detection.

Explication aims at making the user understand how the system behaves.

Justification is about finding an argument why a particular decision is reasonable or normatively acceptable. Justifications are post-hoc.

Part 2: Tasks underlying explainable AI. Some participants suggested that, in order to build a unified conceptual framework for explainable AI, distinguishing the different tasks that fall under the label "explanation in AI" might be more effective than finding satisfactory

definitions of the notions above. We discussed this issue by taking, as a toy example, a system that takes as input a patient's description and returns as output a prediction of whether the patient will be alive in five years. The tasks we identified are as follows:

1. Extract the mechanism that leads from input to output. This task only applies to the case in which the system we are working with is interpretable. The aim underlying the task is to understand how the system produces a prediction. The target (or audience) are the designers of the system. Importantly, the specific form and level of abstraction of the extracted mechanism depends on the designer and what specifically they want to understand. For instance, suppose that the system is based on case-based reasoning but, because of their training, the designer understands abstract argumentation theory better than case-based reasoning. Then the designer might extract the mechanism underlying the system by mapping the case-based reasoning system into abstract argumentation theory.
2. Make the system "user friendly." Once the designer has extracted the mechanism underlying the system, there is a further question of how to make this mechanism accessible to a user. Continuing on the example of a case-based reasoning system, a designer might understand how the system works by proving that the system reaches a decision when, say, the grounded extension of the argumentation framework the system was mapped to contains certain arguments. Of course, things are different for the user, who has probably never heard of abstract argumentation frameworks and grounded semantics. But the designer could make the system "user friendly" by making it capable, first, of extracting an argumentative explanation from the abstract argumentation framework in question and, second, of producing a text containing a translation of the explanation in natural language. As before, the specific form and level of abstraction of the generated explanation depends on the target user and the context.
3. Compare the extracted mechanism with other mechanisms. While tasks 1 and 2 are about understanding how the system works, this task is about finding reasons to accept the predictions of the system. In the toy example of a system that predicts whether a patient will be alive in five years, the task could be understood as comparing the answers to the questions "why did the system predicted that the patient will be alive in five years?" and "why is it reasonable to think that the patient will be alive in five years?". Comparing the answers to the two questions is a way to assess the trustworthiness of the system. In case the system we are working with is a black box and it is not possible to answer the first question, answering the second question is a way to produce a post-hoc justification of the predictions of the system.
4. Extract normative justifications. In the normative domain, justified decisions are decisions that comply with a set of norms. In case the system we are working with is trained on a normative dataset or the mechanism underlying it is subject to normative constraints, then an additional task is to justify the system's decisions by verifing that they were reached without violating the relevant norms.

4.3 Modeling normative reasons

Aleks Knoks (University of Luxembourg, LU), Christoph Benzmüller (Universität Bamberg, DE), Huimin Dong (Sun Yat-Sen University – Zhuhai, CN), Joris Hulstijn (University of Luxembourg, LU), Eric Pacuit (University of Maryland – College Park, US), Antonino Rotolo (University of Bologna, IT), Christian Straßer (Ruhr-Universität Bochum, DE), and Leendert van der Torre (University of Luxembourg, LU)

When philosophers talk about normative matters – about what is right, obligatory, permitted, and so on – they tend to rely on the notion of a normative reason. In the practical domain – which includes morality, as well as the domain of practical rationality – normative reasons are understood as considerations that count in favor of or against actions. The notion has become a mainstay of philosophy, where it is very often relied on in answering various normative and metanormative questions. The so-called reasons-first program takes this to the extreme, taking the notion of a normative reason to be basic and holding that all other normative notions are to be analyzed in terms of it [1, 2, 3]. When discussing the interaction between reasons, the philosophical literature uses such phrases as "the action supported on the balance of reasons" and "reasons for outweigh reasons against", inviting an image of a weighing scale. Philosophers have explored various ideas about the exact workings of this normative weighing scale, with rare exceptions, their investigations have been carried out informally. The overall goal of this breakout session, then, was to model normative reasons and their interaction – roughly, the normative weighing scale for reasons – using methods from mathematical modeling and knowledge representation.

Toward the end of the discussion, the following model emerged:

A structure $T = \langle P, I, F, f, N, D\rangle$ is a model of reasons, where:

- P is a set of persons;
- I is a set of issues;
- F is a set of features;
- $f : P \times I \to 2^F$ is a function mapping pairs of persons and issues to a set of features, indicating which features serve as normative reasons in determining the normative status of the issue for the given person;
- N is a set of polarity functions, with each element n of N having the form $n : P \times I \times \{f\} \times F \to \{+, -, 0\}$ (as their name suggests, these functions determine the polarity or directedness of features);
- D is a set of deontic functions, with each element d of D having the form $d : P \times I \times f \times N \to \{+, -, 0\}$ (as the name suggests, these functions determine the normative status of issues for persons: intuitively, an assignment of $+$ means that the issue (action) is obligatory for the person, that of $-$ means that it is forbidden, and that of 0 means that it is indifferent).

It was noted that, its simplifying assumptions notwithstanding, this model captures important parts of philosophers' way of thinking about normative reasons, the interaction between reasons (or weighing reasons), and the relation between reasons and such normative notions as obligations and permissions. It was also noted that this model is only a starting point, and that more structure can be added to it with ease. For instance, one could make deontic functions depend on additional arguments (representing other normatively relevant information), or one could allow for multiple types of deontic functions (representing different types of obligations). One could also substitute the polarity functions with numerical

functions, with the result that the features identified as reasons for (against) an issue would be associated not only with polarities, but also with magnitudes, bringing the model even closer to the metaphor of weight scales. The affinities with the field of multi-criteria decision-making [4] were also noted.

The discussion participants agreed that the model can be used to formalize various sorts of methodological questions, making them (more) tractable. Questions prompted by the talks delivered at the seminar served as examples: What is the role of polarity (or "directedness") in reason-based decisions? Are Raz's views on reasons in the practical domain equivalent to Pollock's views in the epistemic domain?

For completeness, it should be added that the breakout session participants also voiced some reservations toward the model. Thus, it was noted that the model does not (yet) represent normatively relevant considerations that are not reasons – including conditions and modifiers – which are widely discussed in the philosophical literature. Another issue that was noted was that the model allows one to represent only situations of binary choice. Still, all participants agreed that these issues can be overcome.

References

1 Scanlon T. M. *What We Owe to Each Other*. Harvard University Press, 1998.
2 Raz J. *Practical Reason and Norms*. Oxford University Press, 1990.
3 Parfit D. *On What Matters*, vol. I, Oxford University Press, 2011.
4 Keeney R. and Raiffa H. *Decisions with Multiple Objectives*. Cambridge University Press, 1993.

4.4 Normative reasoning for autonomous agents (Parts I and II)

Pedro Cabalar (University of Coruña, ES), Agata Ciabattoni (Vienna University of Technology, AT), Mehdi Dastani (Utrecht University, NL), Louise A. Dennis (University of Manchester, GB), Huimin Dong (Sun Yat-Sen University – Zhuhai, CN), Thomas Eiter (Vienna University of Technology, AT), Eleonora Giunchiglia (Vienna University of Technology, AT), Guido Governatori (Tarragindi, AU)

From self-driving cars and unmanned aerial vehicles to robot nannies and elder care robots, the myriads of practical uses of Artificial Intelligence (AI) only continue to grow. In these applications an increasingly prominent role is played by autonomous agents, which should operate in an "intelligent" way on some users's behalf but without human intervention. Autonomous agents must accomplish a variety of real world tasks and need to adapt to potentially unpredictable changes in their environment. Reinforcement Learning (RL) – a prominent machine learning technique – has demonstrated to be an effective tool for teaching agents such behaviour [1].

As we assign more roles to RL-based agents it becomes crucial to ensure that they act in ways that are legal, ethics-sensitive, and socially acceptable. This introduces a further challenge: establishing boundaries around the behaviour of these agents, i.e. equipping them with the ability to comply with legal, ethical and social norms, while still enacting pre-learned optimal behaviour.

We have recognized the different approaches that emerge and did thoroughly discuss them. The first approach uses symbolic AI techniques (a.k.a. Logic, Knowledge Representation and Reasoning) and was successfully employed, e.g., in [2] where a theorem proved for a

defeasible deontic logic advises the learning agent on the compliant actions; this approach can however be computationally expensive and less suited for dealing with (signal-based) data, sensory input, or stochastic environments. The other approach relies on sub-symbolic AI (a.k.a. Machine Learning), and was applied, e.g., in [3], to constraint the behaviour of AI agents via reward/penalities; this approach excels under these conditions and enables the construction of efficient and adaptable AI systems, which however lack modularity and transparency; moreover, it is not clear how to adapt this approach to deal with complex normative systems.

All participants agreed that the best way to proceed would be to interlace the two approaches, thus providing the best of both worlds. This breakout session consisted of two parts.

PART I. The participants have identified the main steps for achieving this very challenging task: first translate the norms into efficiently computable representations of normative knowledge, and afterwards to exert some form of normative reasoning to be integrated with the agent's training. Concrete candidates for the norm translations have been proposed: Answer Set Programming [4] and computationally-oriented deontic logics (e.g., Defeasible Deontic Logic [6]).

PART II. (A subgroup[1] of the) participants have concretely discussed potentially useful tools for the second step, and also feasible case studies that could be employed to test their effectiveness and feasibility. There was consensus that the integration of normative reasoning and the Machine Learning component would be the most complex part of the enterprise. To this aim the participants agree that it is worth trying to adapt/extend the techniques used in the Safe Reinforcement Learning community, and/or the emerging idea of constraining Machine Learning with logical formulas [5].

References

1 D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. P. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis. Mastering the game of Go without human knowledge. Nat., 550(7676):354–359, 2017.

2 E. A. Neufeld, E. Bartocci, A. Ciabattoni, and G. Governatori. Enforcing ethical goals over reinforcement-learning policies. J. of Ethics and Inform. Techn., 2022

3 R. Noothigattu, D. Bouneffouf, N. Mattei, R. Chandra, P. Madan, K. R. Varshney, M. Campbell, M. Singh, and F. Rossi. Teaching AI agents ethical values using reinforcement learning and policy orchestration. In Proc. IJCAI, 2019.

4 G. Brewka, T. Eiter, and M. Truszczynski, editors. AI Magazine: special issue on Answer Set Programming. AAAI Press, 2016. Volume 37, number 3. Editorial pp. 5-6.

5 E. Giunchiglia, M. Stoian, T. Lukasiewicz: Deep Learning with Logical Constraints. IJCAI 2022: 5478-5485

6 G. Governatori: Practical Normative Reasoning with Defeasible Deontic Logic. Reasoning Web 2018: 1-25

[1] Some of the participants had moved to other breakout sessions.

4.5 The normative competence of artificial agents

Kevin Baum (University of Saarbrucken, DE), Jan Broersen (Utrecht University, N), Louise A. Dennis (University of Manchester, GB), Frank Dignum (Umeå University, SE), Virginia Dignum (Umeå University, SE), Bertram Malle (Brown University, USA), Xavier Parent (Vienna University of Technology, A), Marija Slavkovik (University of Bergen, NO), Kai Spiekermann (London School of Economics, UK)

This scenario seems to repeat itself: A company creates a service that uses artificial intelligence and/or has some agency. We will refer to this service very generically as "a machine". In its interaction with the users, this machine inevitably violates a norm. The company responds with a constraints that disables the machine from violating the norm, effectively turning the norm into a constraint. Then a new situation arises in which the machines constrained behaviour violates another norm. This practice does not result with a machines ability to operate in a normative context, but rather with a system whose behaviour is neither desirable, nor predictable. To be able to move away from this trap of update and adjust, we first need to ensure there is an understanding on what instruments are available for adjusting and guiding the behaviour of machines.

The work group discussed the basic concepts in normatively regulating the behaviour of machines and artificial agents, as well as the basic approaches. machine:

- Functional level: the machine is either constrained to operating in an environment in which norm violation cannot happen. Of course unintended norm violations can never be ruled out entirely.
- Normative level: the machine is provided with norms that reduce the action space to those actions that comply with the norms (most likely, most of the time).
- Value level: the machine is provided with values with which it needs do align its choice of norm-guided actions, especially when norm conflicts arise.

What intervention we choose to do depends on many factors. The aim of the group is provide a joint article that can be used as an interface to the state of the art in the field of normative reasoning withing multi-agent systems.

5 Open problems

5.1 Is HOL (as a metalogic) all we need for flexible normative reasoning?

Christoph Benzmüller (Universität Bamberg, DE)

Joint work of Christoph Benzmüller, Xavier Parent, Leendert W. N. van der Torre, David Fuenmayor, Aleaxander Steen, Geoff Sutcliffe
Main reference Christoph Benzmüller, Xavier Parent, Leendert W. N. van der Torre: "Designing normative theories for ethical and legal reasoning: LogiKEy framework, methodology, and tool support", Artif. Intell., Vol. 287, p. 103348, 2020.
URL https://doi.org//10.1016/j.artint.2020.103348
Main reference Christoph Benzmüller, David Fuenmayor, Alexander Steen, Geoff Sutcliffe: "Who Finds the Short Proof?", Logic Journal of the IGPL, p. jzac082, 2023.
URL https://doi.org/10.1093/jigpal/jzac082

In previous work we have shown that classical higher-order logic (HOL), when used as a metalogic, enables (shallow) semantic embeddings of various state-of-the-art logics for normative reasoning. To this end, the logico-pluralistic LogiKEy [1] methodology and framework has been developed to support both metalogical studies of logics for normative reasoning [2] and their applications [3].

In this talk I summarise these developments and ask the obvious question: Is HOL already all we need to support flexible normative reasoning on computers? Or are there logics for normative reasoning that cannot be addressed by the LogiKEy approach?

We also briefly address typical arguments against HOL, namely that undecidability and complexity considerations militate against its use. With reference to very recent practical work on speeding up proofs in HOL [4], we will take a partially contrary position.

References

1 C. Benzmüller, X. Parent, L. van der Torre. Designing Normative Theories for Ethical and Legal Reasoning: LogiKEy Framework, Methodology, and Tool Support. Artificial Intelligence, 287: 103348. 2020. http://doi.org/10.1016/j.artint.2020.103348 (Preprint: https://www.researchgate.net/publication/342146653)

2 X. Parent, C. Benzmüller. Automated Verification of Deontic Correspondences in Isabelle/HOL – First Results. In Benzmüller, C., & Otten, J., editor(s), ARQNL 2022: Automated Reasoning in Quantified Non-Classical Logics. Proceedings of the 4th International Workshop on Automated Reasoning in Quantified Non-Classical Logics (ARQNL 2022) affiliated with the 11th International Joint Conference on Automated Reasoning (IJCAR 2022). Haifa, Israel, August 11, 2022, volume 3326, pages 92-108, 2023. CEUR Workshop Proceedings, CEUR-WS.org. https://ceur-ws.org/Vol-3326/

3 D. Fuenmayor, C. Benzmüller. Normative Reasoning with Expressive Logic Combinations. In De Giacomo, G., Catala, A., Dilkina, B., Milano, M., Barro, S., Bugarín, A., & Lang, J., editor(s), ECAI 2020 – 24th European Conference on Artificial Intelligence, June 8-12, Santiago de Compostela, Spain, volume 325, of Frontiers in Artificial Intelligence and Applications, pages 2903-2904, 2020. IOS Press. http://doi.org/10.3233/FAIA200445

4 C. Benzmüller, D. Fuenmayor, A. Steen, G. Sutcliffe. Who Finds the Short Proof? Logic Journal of the IGPL. 2023. http://doi.org/10.1093/jigpal/jzac082 (Preprint: https://www.researchgate.net/publication/367464450)

Participants

- Guillaume Aucher
University of Rennes, FR
- Kevin Baum
DFKI – Saarbrücken, DE
- Christoph Benzmüller
Universität Bamberg, DE
- Jan M. Broersen
Utrecht University, NL
- Pedro Cabalar
University of Coruña, ES
- Ilaria Canavotto
University of Maryland – College Park, US
- Agata Ciabattoni
TU Wien, AT
- Célia da Costa Pereira
Université Côte d'Azur – Sophia Antipolis, FR
- Mehdi Dastani
Utrecht University, NL
- Louise A. Dennis
University of Manchester, GB
- Frank Dignum
University of Umeå, SE
- Virginia Dignum
University of Umeå, SE
- Huimin Dong
Sun Yat-Sen University – Zhuhai, CN
- Thomas Eiter
TU Wien, AT
- Eleonora Giunchiglia
TU Wien, AT
- Guido Governatori
Tarragindi, AU
- John F. Horty
University of Maryland – College Park, US
- Joris Hulstijn
University of Luxembourg, LU
- Aleks Knoks
University of Luxembourg, LU
- Emiliano Lorini
CNRS – Toulouse, FR
- Bertram F. Malle
Brown University – Providence, US
- Réka Markovich
University of Luxembourg, LU
- Eric Pacuit
University of Maryland – College Park, US
- Xavier Parent
TU Wien, AT
- Bijan Parsia
University of Manchester, GB
- Adrian Paschke
FU Berlin, DE
- Henry Prakken
Utrecht University, NL
- Antonino Rotolo
University of Bologna, IT
- Ken Satoh
National Institute of Informatics – Tokyo, JP
- Marija Slavkovik
University of Bergen, NO
- Kai Spiekermann
London School of Economics, GB
- Christian Straßer
Ruhr-Universität Bochum, DE
- Leon van der Torre
University of Luxembourg, LU