Formal Methods for Responsibility Reasoning in Multiagent Systems

Vahid Yazdanpanah[1] and Mehdi Dastani[2]

[1]University of Southampton, United Kingdom
[2]Utrecht University, The Netherlands

April 2021

Abstract. Safe and reliable deployment of collaborative AI-human multiagent systems requires formal semantics and verifiable tools to reason about forward-looking responsibilities of agents (e.g., who can/should ensure some properties in prospect) as well as their backward-looking responsibilities (e.g., who to blame, praise, or see accountable for an already materialized outcome in retrospect) [3, 2, 1]. Modeling and reasoning about responsibility calls for capturing its *strategic*, *epistemic*, and *normative* aspects for which the community of formal methods possesses apt semantic systems and reasoning tools. In this talk, we report on a line of research on the application of formal methods and modal logics for reasoning about different forms of responsibility in multiagent systems [4, 5, 6]. In addition, we overview recent work on the application of responsibility reasoning for task coordination, discuss open problems [7], and highlight the potentials of formal responsibility reasoning in AI systems.

Acknowledgements. This work is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) through the Trustworthy Autonomous Systems Hub (EP/V00784X/1) and the platform grant entitled “AutoTrust: Designing a Human-Centred Trusted, Secure, Intelligent and Usable Internet of Vehicles” (EP/R029563/1).

References

[1] Natasha Alechina, Joseph Y. Halpern, and Brian Logan. Causality, responsibility and blame in team plans. In *In Proceedings of AAMAS-2017*, pages 1091–1099, 2017.

[2] Hana Chockler and Joseph Y. Halpern. Responsibility and blame: A structural-model approach. *J. Artif. Intell. Res.*, 22:93–115, 2004.

[3] Ibo van de Poel. The relation between forward-looking and backward-looking responsibility. In *Moral responsibility*, pages 37–52. Springer, 2011.

[4] Vahid Yazdanpanah and Mehdi Dastani. Quantified degrees of group responsibility. In *Proceedings of the International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems*, pages 418–436. Springer, 2015.

[5] Vahid Yazdanpanah and Mehdi Dastani. Distant group responsibility in multi-agent systems. In *Proceedings of the International Conference on Principles and Practice of Multi-Agent Systems*, pages 261–278. Springer, 2016.

[6] Vahid Yazdanpanah, Mehdi Dastani, Wojciech Jamroga, Natasha Alechina, and Brian Logan. Strategic responsibility under imperfect information. In *Proceedings of AAMAS-2019*, pages 592–600, 2019.

[7] Vahid Yazdanpanah, Enrico H. Gerding, Sebastian Stein, Mehdi Dastani, Catholijn M. Jonker, and Timothy J. Norman. Responsibility research for trustworthy autonomous systems. In *Proceedings of AAMAS-2021*, page 57–62, 2021.