▸ EKATERINA KUBYSHKINA, GIUSEPPE PRIMIERO, *Trustworthy AI: probabilities meet possible worlds.*
Logic, Uncertainty, Computation, Information Group, Department of Philosophy, University of Milan, Via Festa del Perdono 7, 20122 Milan, Italy.
E-mail: ekaterina.kubyshkina@unimi.it.
Logic, Uncertainty, Computation, Information Group, Department of Philosophy, University of Milan, Via Festa del Perdono 7, 20122 Milan, Italy.
E-mail: giuseppe.primiero@unimi.it.

The notion of trustworthiness, central to many fields of human inquiry, has recently attracted the attention of various researchers in logic, computer science and artificial intelligence (AI). Both conceptual and formal approaches for modelling trustworthiness as a (desirable) property of AI systems are emerging in the literature. To develop logics fit for this aim means to analyse both the non-deterministic aspect of AI systems and to offer a formalization of the intended meaning of their trustworthiness. In this work we take a semantic perspective on representing such processes, and provide a measure on possible worlds for evaluating them as trustworthy. In particular, we intend trustworthiness as the correspondence within acceptable limits between a model in which the theoretical probability of a process to produce a given output is expressed and a model in which the frequency of showing such output as established during a relevant number of tests is measured. From a technical perspective, we show that our semantics characterizes the probabilistic typed natural deduction calculus introduced in [1] and further extended in [2]. This contribution connects those results on trustworthy probabilistic processes with the mainstream method in modal logic, thereby facilitating the understanding of this field of research for a larger audience of logicians, as well as setting the stage for an epistemic logic appropriate to the task.

[1] Fabio Aurelio D'Asaro and Giuseppe Primiero, *Probabilistic typed natural deduction for trustworthy computations*, ***Proceedings of the 22nd International Workshop on Trust in Agent Societies (TRUST 2021) Co-located with the 20th International Conferences on Autonomous Agents and Multiagent Systems (AAMAS)*** (London, UK), (Dongxia Wang, Rino Falcone, and Jie Zhang,), vol. 3022, CEUR Workshop Proceedings, 2021.

[2] Fabio Aurelio D'Asaro, Francesco Genco and Giuseppe Primiero, *Checking trustworthiness of probabilistic computations in a typed natural deduction system*, ***CoRR***, abs/2206.12934, 2022.