

1. Record Nr.	UNINA9910882900103321
Autore	Osman Nardine
Titolo	Value Engineering in Artificial Intelligence : First International Workshop, VALE 2023, Krakow, Poland, September 30, 2023, Proceedings // edited by Nardine Osman, Luc Steels
Pubbl/distr/stampa	Cham : , : Springer Nature Switzerland : , : Imprint : Springer, , 2024
ISBN	3-031-58202-0
Edizione	[1st ed. 2024.]
Descrizione fisica	1 online resource (272 pages)
Collana	Lecture Notes in Artificial Intelligence, , 2945-9141 ; ; 14520
Altri autori (Persone)	SteelsLuc
Disciplina	006.3
Soggetti	Artificial intelligence Artificial Intelligence
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Frameworks for Norms and Values: That's All Folks: a KG of Values as Commonsense Social Norms and Behaviors -- Towards a Formalisation of Motivated Reasoning and the Roots of Conflict -- Perspective-Dependent Value Alignment of Norms. Detection of Moral Values: Moral values in social media for disinformation and hate speech analysis -- Social Value Alignment in Large Language Models -- Do Language Models Understand Morality? Towards a Robust Detection of Moral Content -- Detection and Analysis of Moral Values in Argumentation. Learning and Engineering of Policies: Algorithms for learning value-aligned policies considering admissibility relaxation -- On Autonomy, Governance, and Values: An AGV Approach to Value Engineering -- Exploiting value system structure for value-aligned decision-making -- On value-aligned cooperative multi-agent task allocation. Implementation of Norms: Values, proportionality, and uncertainty in military autonomous devices -- Towards a Distributed Platform for Normative Reasoning and Value Alignment in Multi-Agent Systems -- Value-based Reasoning Scenario in Employee Hiring and Onboarding using Answer Set Programming -- Value awareness and process automation: a reflection through school place allocation models.
Sommario/riassunto	This book constitutes the post-conference proceedings of the First International Workshop on Value Engineering in Artificial Intelligence, VALE 2023, held in Krakow, Poland, on September 30, 2023. The 16

full papers presented were carefully reviewed and selected. The papers have been grouped into four themes, addressing a number of research questions for each:

- I. Frameworks for norms and values: What are existing or new frameworks for norms and values and how can they be represented in AI systems in order to make these systems norm- and value-aware?
- II. Detection of moral values: Given an existing AI system, how can we figure out whether its behavior is aligned with human values?
- III. Learning and engineering of value-aligned policies: How can policies for implementing norms and values be acquired or designed?
- IV. Implementation of norms: How can norms be implemented in concrete AI applications?
