

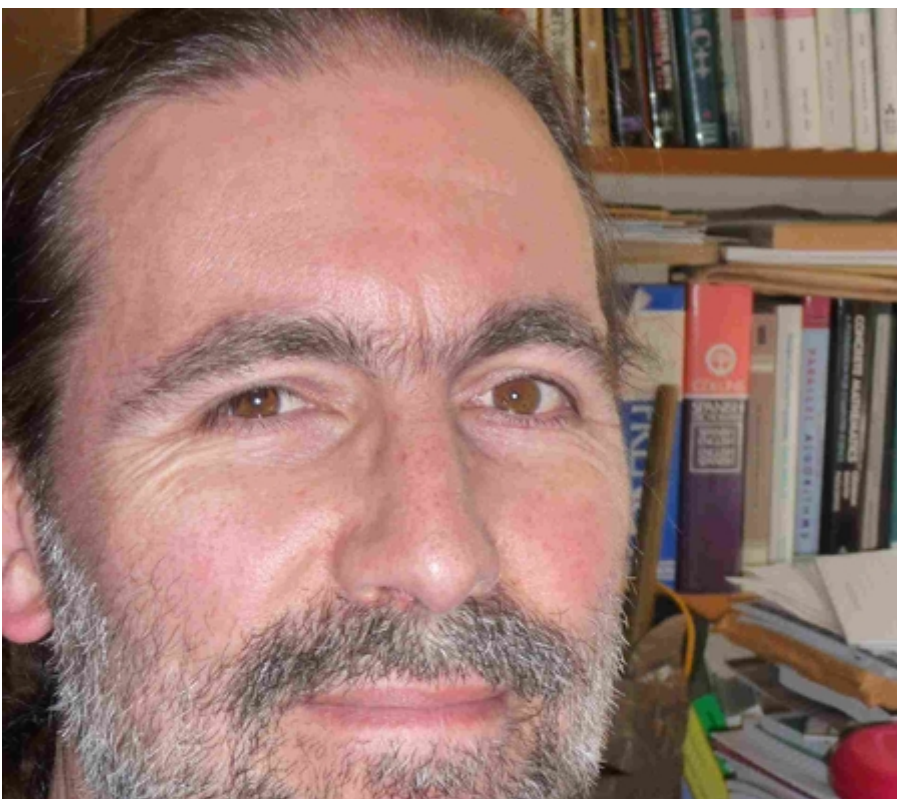
[Inici](#) > Hybrid BSC RS: An introduction to conscientious design and an exploration of the operationalization of values for digital twins and other hybrid social systems

Hybrid BSC RS: An introduction to conscientious design and an exploration of the operationalization of values for digital twins and other hybrid social systems

Objectives

Abstract: This talk is part software engineering, part artificial intelligence. It is based on two recent papers in conjunction with Mark d'Inverno (Goldsmiths, University of London), Pablo Noriega (IIIA), and Harko Verhagn (Stockholm University): Ethical Online AI Systems Through Conscientious Design (<https://doi.org/10.1109/MIC.2021.3098324>) and Design Heuristics for Ethical Online Institutions (<https://coin-workshop.github.io/coine-2022-auckland/papers/paper-12.pdf>).

The goal of Conscientious Design (CD) is to enable embedding stakeholder values in online environments, where software and humans interact across the cyber-physical divide. The problem is how to do this in a structured, repeatable way, and how to do it so that stakeholders are co-owners of the system in control of its evolution in line with changing stakeholder priorities (value preference drift) over its lifetime. CD draws on Schwartz's universal values, Deming's Total Quality Management, Friedman's Value Sensitive Design and Alexander's Timeless way of building to provide conceptual tools to address this problem, along with some preliminary heuristics to support the process of identifying, refining and operationalizing values. We do not claim this is (yet) a solution to the problem of building systems that respond in a way that is consistent with human values - it is very much a work in progress - but it does begin to provide a structured way to dissect this problem space. As such, this talk is as much seeking feedback as it is exposing ideas.



Julian Padget is a Reader in AI in the Department of Computer Science at the University of Bath. His research interests are in intelligent agents, norm representation and reasoning, and distributed systems. He began work on how to represent software-interpretable constraints on behaviour over 20 years ago, publishing formal models for distributed auctions. Since then, a common thread throughout his work has been how to translate human requirements into verifiable machine-processable descriptions, such as current work on data security policies, legal reasoning, verification of (declarative) smart contracts (for distributed ledgers), and the evolution of machine-processable governance policies in autonomous systems. He is active in the standards community, where he participates in several AI-related committees: ISO/SC42/WG2 (Data), WG3 (Trustworthiness) and WG5 (AI systems), CEN/JTC21 (AI), IEEE P7003 (Consideration of algorithmic bias). He is a member of the UN Committee of Experts on Big Data Privacy Preserving Techniques (PPTs) task team and the Legal Aspects of PPTs task team.

Speakers

Speaker: Julian Padget is a Reader in AI in the Department of Computer Science at the University of Bath.

Host: Ulises Cortés, High Performance Artificial Intelligence Group Manager, BSC

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 14 jul 2024 - 09:49): <https://www.bsc.es/ca/research-and-development/research-seminars/hybrid-bsc-rs-introduction-conscientious-design-and-exploration-the-operationalization-values>