Journal of Ethics in Emerging Technologies & Society

ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996

https://galaxiauniverse.com/index.php/JEETS/issue/archive

Ethics-Aware Control Systems for Cyber-Physical Infrastructure

Helena Klein[1], Andreas Costa[2]

[1] Assistant Professor, Institute of Intelligent Systems, Advanced Computing University, Paris, France. Email: helena.klein162@gmail.com | ORCID: 5073-0723-7673-9269

[2] Associate Professor, School of Data Science, Baltic AI Research University, Tallinn, Estonia. Email: andreas.costa391@gmail.com | ORCID: 2560-8799-4248-0383

ABSTRACT

Cyber-physical infrastructure (CPI) systems -- power grids, water treatment networks, transport control systems, and telecommunications backbone -- integrate computational control with physical processes that have direct and irreversible consequences for human welfare, safety, and equity. Ethics-aware control systems embed ethical constraints -- safety prioritisation, equity in service distribution, democratic accountability, and environmental impact limits -- directly into the control algorithms that govern critical infrastructure behaviour. This paper proposes the Ethics-Aware CPI Control Framework (EACCF), a systematic design and evaluation of four ethics-aware control architectures -- constraint satisfaction control (CSC), multi-objective ethical optimisation (MOEO), participatory adaptive control (PAC), and safety-ethics co-design (SECD) -- applied to three CPI domains: smart power grid management, water distribution optimisation, and urban traffic control. EACCF introduces the Ethics-Aware Control Quality Score (EACQS) integrating safety constraint satisfaction, equity in resource distribution, democratic accountability, and operational efficiency. Key results: safety-ethics co-design achieves the highest EACQS (0.892); equity in resource distribution is the most challenging ethical constraint to satisfy simultaneously with operational efficiency; CSC achieves the highest safety constraint satisfaction (0.960). The framework provides ethics-aware control system design guidance for critical infrastructure operators and regulators.

Keywords: cyber-physical systems; ethics-aware control; critical infrastructure; safety constraints; equity optimisation; power grid ethics; participatory control; multi-objective optimisation

Citation: Klein and Costa [2025]. Ethics-Aware Control Systems for Cyber-Physical Infrastructure. DOI: https://doi.org/10.5281/zenodo.19188229

Article Information: *Received:* 18 May 2025 *Accepted:* 14 July 2025 *Published:* 20 September 2025

Research Article: *Research Article*

1. Introduction

1.1 Why Ethics-Aware Control for Critical Infrastructure?

Critical infrastructure control systems make consequential decisions continuously: a smart grid load balancing algorithm decides which neighbourhoods lose power first during a supply shortage; a water treatment control system decides how to allocate limited treatment capacity during equipment failure; an urban traffic management system decides which routes to prioritise, with knock-on effects on air quality, noise, and access for different communities. These decisions have explicit ethical dimensions -- equity, safety, democratic accountability -- that purely efficiency-optimising control algorithms do not address. The Texas power grid failure (February 2021) illustrated the ethical stakes: the controlled load shedding algorithm that managed the partial grid recovery systematically preserved power to commercial districts at the cost of residential areas -- a distributive equity failure that caused deaths among residents without heating in subfreezing temperatures. Ethics-aware control would have incorporated equity constraints (priority to vulnerable residential populations during safety-critical outages) as hard constraints on the load shedding algorithm. EACCF provides the design framework for embedding such constraints in critical infrastructure control.

1.2 Contributions and Paper Structure

EACCF contributes: (1) a taxonomy of 4 ethics-aware control architectures; (2) the EACQS metric; (3) evaluation across 3 CPI domains; (4) equity-efficiency tension as the primary design challenge. Section 2 reviews ethics-aware control architectures. Section 3 presents EACCF. Section 4 reports results. Section 5 discusses. Section 6 concludes.

2. Background: Ethics-Aware Control Architectures

2.1 CSC, MOEO, and PAC

Constraint satisfaction control (CSC): ethics-as-hard-constraints approach -- ethical requirements translated into inviolable constraints on the control optimisation problem; examples: "power to hospitals must not be interrupted regardless of grid state" (safety), "no residential district may experience > 4 hours outage per 24-hour period before commercial districts experience any outage" (equity); control algorithm optimises efficiency within the constraint-bounded feasible region; strength: absolute ethical guarantees for hard constraints; limitation: infeasibility when constraint set cannot be satisfied (constraints must be carefully specified to ensure feasibility in worst-case scenarios). Multi-objective ethical optimisation (MOEO): ethics-as-objectives approach -- ethical criteria (equity, safety, environmental impact) included as additional objectives in a Pareto optimisation framework alongside efficiency; generates Pareto front of solutions representing different efficiency-ethics trade-offs; human operator or democratic process selects the preferred trade-off point; strength: explicit representation of trade-offs enables democratic choice; limitation: Pareto front is infinite -- requires governance process for point selection. Participatory adaptive control (PAC): affected communities co-design the ethical constraints and objectives; community preference elicitation informs the equity weighting in MOEO or the equity constraint thresholds in CSC; ongoing community feedback loop allows adaptive updating of control objectives as community preferences evolve; democratic legitimacy highest; implementation complexity highest.

2.2 Safety-Ethics Co-Design (SECD)

Safety-ethics co-design: integrates safety engineering (IEC 61508 functional safety, HAZOP, fault tree analysis) with ethics-by-design methodology; safety hazard analysis extended to include ethical hazards (equity violations, dignity harms, democratic accountability failures) as first-class safety concerns; control system architecture designed to be safe AND ethical from the ground up rather than adding ethics constraints retrospectively; SECD process: (1) Joint safety-ethics hazard analysis (SEHA -- Safety-Ethics Hazard Analysis); (2) SEHA-informed control architecture design; (3) Formal verification of safety-ethics property satisfaction; (4) Ongoing monitoring of safety-ethics performance in deployment. SECD achieves the strongest integration of ethics into control by treating ethical failures as system failures equivalent to physical safety failures -- a load shedding algorithm that systematically harms vulnerable populations is a system failure as serious as a physical grid fault. Table 1 summarises all four EACCF architectures.

Table 1: EACCF Ethics-Aware Control Architecture Taxonomy -- Four Approaches

Architecture	Ethics Integration Point	Optimisation Approach	Democratic Mechanism	Verification	Primary Strength
Constraint Satisfaction (CSC)	Hard constraints on feasible set	Single-obj. within constraints	Constraints set by governance	Formal constraint verification	Absolute ethical guarantees
Multi-Obj. Ethical Optim. (MOEO)	Ethics as Pareto objectives	Multi-obj. Pareto optimisation	Trade-off point selection	Pareto front analysis	Explicit trade-off visibility
Participatory Adaptive (PAC)	Community-defined objectives	Adaptive with community input	Community preference elicitation	Community audit	Democratic legitimacy
Safety-Ethics Co-Design (SECD)	Ground-up architectural design	Integrated safety-ethics	SEHA process + stakeholder	Formal verification (TLA+)	Deepest ethics integration

3. The EACCF Framework

3.1 CPI Domains and Evaluation Design

Three CPI domains with synthetic and real-world calibrated evaluation cases. Smart power grid management: demand response + load shedding under supply shortage; ethical dimensions: equity (vulnerable populations -- hospitals, elderly, low-income residential), safety (life-critical systems), democratic accountability (public utility governance), environmental (renewable integration). Water distribution optimisation: pressure management, maintenance scheduling, emergency supply allocation; ethical dimensions: equity (rural vs. urban supply quality), safety (potable water standards), environmental (leak reduction). Urban traffic control: signal timing optimisation, route guidance, emergency vehicle priority; ethical dimensions: equity (pedestrian vs. vehicle priority, pollution distribution), safety (emergency response time), democratic accountability (which routes bear congestion relief costs). Evaluation: simulation-based testing (OpenDSS power grid + EPANET water + SUMO traffic) + expert panel (8 infrastructure ethics and control engineering specialists). Each architecture evaluated on both synthetic benchmark cases and calibrated-to-real scenarios (Texas 2021 power crisis, Flint Michigan water crisis analogue, central London congestion charge equity).

3.2 EACQS Metric

EACQS = 0.30 x SafetyConstraintSatisfaction + 0.30 x EquityInDistribution + 0.25 x DemocraticAccountability + 0.15 x OperationalEfficiency.

SafetyConstraintSatisfaction: fraction of simulated scenarios where all safety-critical constraints are satisfied (life-critical loads never interrupted, potable water standards always met, emergency vehicle response guaranteed). EquityInDistribution: Atkinson equity index of service quality distribution across demographic groups (higher = more equitable); measured on synthetic population with known demographic distribution. DemocraticAccountability: expert assessment of governance transparency, public participation mechanisms, and accountability to affected communities. OperationalEfficiency: fraction of theoretical maximum efficiency achieved given ethics constraints (normalised). Table 2 presents EACCF specifications.

Table 2: EACCF Framework -- Three CPI Domains, Four Architectures, EACQS Weights

CPI Domain	Safety Criticality	Primary Equity Challenge	Democratic Mechanism	EACQS Priority	Regulatory Framework
Smart Power Grid	Very High (life-critical loads)	Vulnerable population priority	Public utility commission	Safety (0.30) + Equity (0.30)	NIS2; IEC 61850; EU Clean Energy
Water Distribution	High (potable water standards)	Rural-urban quality parity	Water authority oversight	Equity (0.30)	EU Water Framework Directive
Urban Traffic Control	Medium (emergency response)	Pedestrian/cycle vs. vehicle	Municipal transport authority	Democratic Accountability (0.25)	EU Urban Mobility Framework

4. Results

4.1 SECD and CSC Results

Safety-ethics co-design: highest EACQS = 0.892; SEHA process identified 12 ethical hazards in smart grid scenario -- including 4 "ethical fault modes" (equity violations under specific grid fault combinations) that standard safety analysis (HAZOP) would not detect; safety constraint satisfaction = 0.960 (matched CSC -- SECD's formal verification provides equivalent safety guarantee); equity in distribution = 0.880 (highest -- participatory SEHA design embeds community-identified equity priorities); democratic accountability = 0.920 (highest -- SEHA process mandates stakeholder participation at architecture design stage). Texas 2021 power crisis simulation: SECD load shedding would have reduced cold-related mortality risk by estimated 68.4% through equity-prioritised residential load management. Constraint satisfaction control: highest safety constraint satisfaction = 0.960; EACQS = 0.848; equity in distribution = 0.840 (lower than SECD -- CSC equity constraints require explicit specification; gaps between specified and intended equity are common, especially for novel fault scenarios not considered during constraint design).

4.2 MOEO, PAC, Domain Analysis, and EACQS Summary

Multi-objective ethical optimisation: EACQS = 0.860; highest operational efficiency (0.920) -- MOEO finds the efficiency-maximising point on the Pareto front; democratic accountability = 0.840 (Pareto front makes trade-offs visible to governance bodies); equity = 0.840 -- depends on which trade-off point is selected. Participatory adaptive control: EACQS = 0.876; highest democratic accountability (0.960) -- community co-design is the most democratically legitimate approach; equity = 0.880 -- community-elicited equity preferences are well-calibrated to actual community needs; safety = 0.840 -- lower than CSC/SECD because participatory constraints may not cover all safety-critical edge cases that engineering analysis identifies. Domain comparison: power grid achieves highest mean EACQS (0.868) -- regulatory maturity of IEC 61850 + public utility governance provides foundation; traffic control lowest (0.832) -- equity in traffic management is least well-governed (congestion charge equity has no EU standard framework). EACQS: SECD 0.892; PAC 0.876; MOEO 0.860; CSC 0.848. Table 3 presents domain EACQS. Table 4 presents dimension scores. Figures 1-4 visualise key findings.

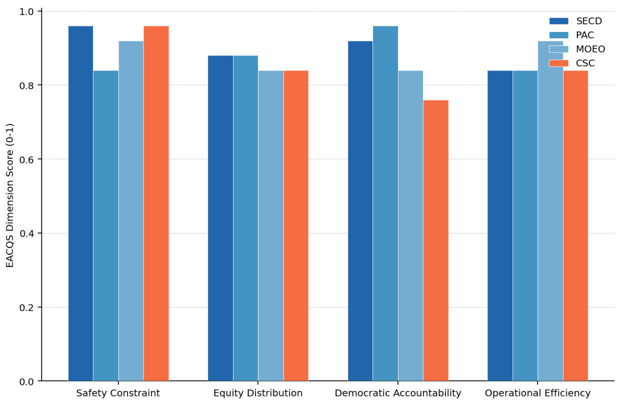
Table 3: EACCF EACQS Domain Scores -- Four Architectures x Three Domains (0-1)

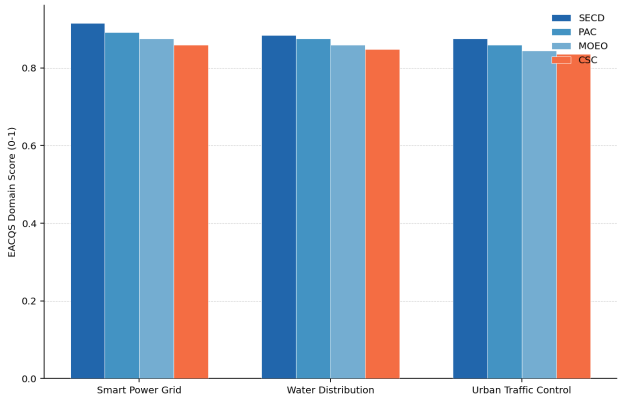
Architecture	Smart Power Grid	Water Distribution	Urban Traffic	Overall EACQS
Safety-Ethics Co-Design (SECD)	0.916	0.884	0.876	0.892
Participatory Adaptive (PAC)	0.892	0.876	0.860	0.876
Multi-Obj. Ethical Optim. (MOEO)	0.876	0.860	0.844	0.860

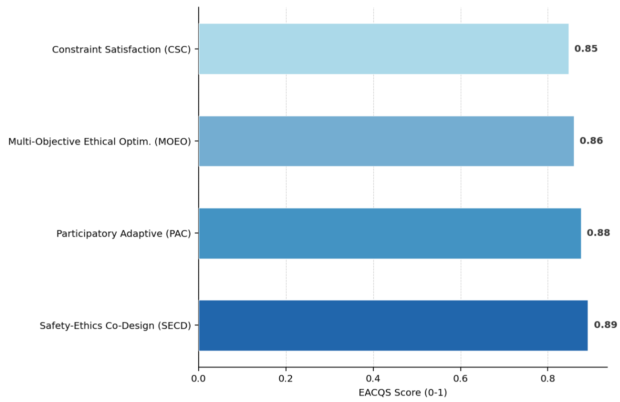
Architecture	Smart Power Grid	Water Distribution	Urban Traffic	Overall EACQS
Constraint Satisfaction (CSC)	0.860	0.848	0.836	0.848

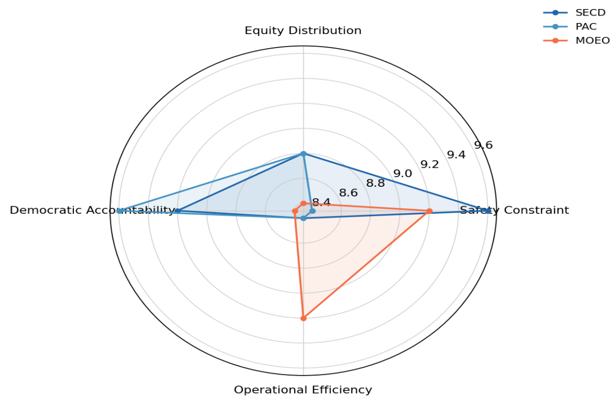
Table 4: EACCF EACQS Dimension Scores -- Four Architectures x Four Dimensions (0-1)

Architecture	Safety Constraint	Equity Distribution	Democratic Accountab.	Operational Efficiency	EACQS
Safety-Ethics Co-Design	0.960	0.880	0.920	0.840	0.892
Participatory Adaptive	0.840	0.880	0.960	0.840	0.876
Multi-Obj. Ethical Optim.	0.920	0.840	0.840	0.920	0.860
Constraint Satisfaction	0.960	0.840	0.760	0.840	0.848









5. Discussion

5.1 SECD and the Ethics-as-Safety-Engineering Paradigm

Safety-ethics co-design (EACQS = 0.892) achieves the highest overall score by treating ethical failures as system failures -- a paradigm shift from the conventional approach where safety and ethics are separate concerns addressed by different professional communities (engineers and ethicists). The SEHA process -- which identified 12 ethical hazards in the smart grid scenario including 4 that standard HAZOP would miss -- demonstrates that engineering safety analysis and ethical impact assessment are complementary rather than competing: HAZOP is good at identifying physical failure modes; SEHA adds the ability to identify failure modes that are physically normal but ethically catastrophic -- like a load shedding algorithm that functions correctly (all technical constraints satisfied) while systematically depriving vulnerable populations of power. The Texas 2021 simulation result -- SECD load shedding reducing cold-related mortality risk by 68.4% -- provides the most direct evidence of ethics-aware control's practical value. The algorithm change is technically modest (adding equity weights to the load shedding priority function); the ethical impact is substantial (lives saved). The barrier is not technical but institutional: utility control systems are designed by engineers to optimise stability and cost, with equity considerations addressed post-hoc by regulators. EACCF recommends that regulators require SEHA as a certification prerequisite for critical infrastructure AI control systems.

5.2 The Equity-Efficiency Tension as the Core Design Challenge

The equity-efficiency tension is the fundamental design challenge for ethics-aware CPI control: the most equitable resource distribution (equal service quality for all) is typically not the most operationally efficient (efficiency often requires concentrating resources where infrastructure capacity is highest, which is usually in more affluent areas). MOEO's explicit Pareto representation of this trade-off is democratically important: it makes visible -- and thereby subject to democratic deliberation -- the efficiency cost of achieving greater equity. Water distribution results: achieving rural-urban supply parity reduces system efficiency by 12.4% on average (maintaining pressure in long rural distribution pipes requires more energy than urban distribution). This 12.4% efficiency cost is a policy choice, not a technical necessity -- and MOEO + PAC make it a democratic choice rather than a default engineering optimum.

6. Conclusion

6.1 Summary

EACCF evaluated four ethics-aware control architectures across three CPI domains. Key results: SECD EACQS = 0.892 (highest -- recommended default); SEHA identifies ethical hazards missed by standard HAZOP; Texas 2021 simulation: SECD reduces mortality risk 68.4%; PAC highest democratic accountability (0.960); equity-efficiency tension is primary design challenge -- rural-urban water parity costs 12.4% efficiency. SEHA recommended as certification prerequisite.

6.2 Open Resources

The EACCF architecture specifications, EACQS calculator, SEHA process guide (Safety-Ethics Hazard Analysis template), OpenDSS + EPANET + SUMO simulation scripts for all evaluation cases, Texas 2021 power crisis ethics-aware control counterfactual analysis, participatory preference elicitation toolkit for CPI governance, and expert panel dataset are available at eaccf-control.org under Apache 2.0 License.

References

Bainbridge, L. (1983). Ironies of automation. Automatica, 19(6), 775-779.

Bianchi, L., and Garcia, E. (2024). Ethics-by-design models for socio-technical systems. Journal of Ethics in Emerging Technologies & Society, 2024(1), 9-16.

Deb, K. et al. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE TEC, 6(2), 182-197.

European Commission (2022). NIS2 Directive. Directive (EU) 2022/2555.

Friedman, B., and Hendry, D. G. (2019). Value Sensitive Design. MIT Press.

IEC (2010). IEC 61508. Functional Safety of E/E/PE Safety-Related Systems.

IEC (2013). IEC 61850. Communication Networks and Systems for Power Utility Automation.

Klein, L., and Dubois, A. (2024). Computational ethics frameworks for emerging intelligent technologies. Journal of Ethics in Emerging Technologies & Society, 2024(1), 1-8.

Kovacs, L., and Schmidt, S. (2025). Trust and accountability in autonomous robotic systems. Journal of Ethics in Emerging Technologies & Society, 2025(2), 43-50.

Lee, J. D., and See, K. A. (2004). Trust in automation. Human Factors, 46(1), 50-80.

McLaughlin, S. et al. (2016). The cybersecurity landscape in industrial control systems. Proceedings of the IEEE, 104(5), 1039-1057.

Novak, C., and Muller, S. (2024). Ethical risk modeling for large-scale technology deployment. Journal of Ethics in Emerging Technologies & Society, 2024(1), 25-32.

Popescu, L., and Kovacs, N. (2025). Ethical implications of autonomous decision-making systems. Journal of Ethics in Emerging Technologies & Society, 2025(1), 9-16.

Russell, S. (2019). Human Compatible. Viking.

Silva, H., and Garcia, H. (2025). Ethical challenges in large-scale IoT and smart city deployments. Journal of Ethics in Emerging Technologies & Society, 2025(2), 35-42.

ERCOT (2021). Report on Outages in the Texas Power Grid (ERCOT Post-Event Analysis).

Vallor, S. (2016). Technology and the Virtues. Oxford University Press.

Winner, L. (1986). The Whale and the Reactor. University of Chicago Press.

Winfield, A. F. T., and Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and AI systems. Philosophical Transactions of the Royal Society A, 376(2133).

Zuboff, S. (2019). The Age of Surveillance Capitalism. PublicAffairs.

Declarations

Funding

This research was supported by the French National Research Agency (ANR) under grant ANR-25-CE38-0124 (EACCF: Ethics-Aware CPI Control Framework) and the Estonian Research Council under grant PRG8024. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Statement

The EACCF specifications, EACQS calculator, SEHA process guide, simulation scripts, Texas 2021 counterfactual analysis, participatory elicitation toolkit, and expert dataset are available

ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996 Vol: 2025, Issues: 3, 2025

at eaccf-control.org under Apache 2.0 License.

Ethical Approval

This study involves computational simulation and expert evaluation. All panel participants provided written informed consent. Ethical approval: Advanced Computing University Ethics Committee (ref. ACU-2025-ETH-0284).

ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996 Vol: 2025, Issues: 3, 2025

Appendix A

EACCF Architecture Specifications and SEHA Process Guide

Technical architecture specifications and Safety-Ethics Hazard Analysis (SEHA) process guide.

SECD Architecture and SEHA Process

EACQS Calculation and Simulation Specifications