

Beyond Optimization: A Philosophical Framework for Responsible Implementation of Technical Systems

[Author Name]

[University/Institution]

[Email]

13th January 2025

Abstract

The implementation of optimization systems in critical domains has outpaced our philosophical frameworks for ensuring their ethical deployment. This paper introduces the Responsible Optimization Implementation Framework (ROIF), grounding it in value pluralism, distributed moral agency, and dynamic ethics. ROIF addresses the fundamental tension between technical optimization and human values, providing both theoretical foundations and practical mechanisms for ethical system implementation. The framework's primary contribution lies in its integration of philosophical rigor with practical applicability, offering a structured approach to managing the ethical complexity of modern optimization systems.

1 Introduction

Optimization systems increasingly mediate critical decisions in healthcare, finance, and public services (Yeung 2018; Zarsky 2016), raising fundamental questions about the relationship between technical efficiency and human values. Current approaches typically treat ethical considerations as external constraints on optimization processes (mittelstadt2016), leading to implementations that fail to adequately address the complex interplay between

technical systems and human values. This inadequacy stems not from a lack of ethical awareness, but from what Winner (1980) identified as a fundamental misconception of how values operate within socio-technical systems.

The challenge lies in the nature of optimization systems themselves. These systems operate at scales and speeds that render traditional ethical oversight mechanisms ineffective (Citron and Pasquale 2014), create distributed impacts that challenge conventional notions of responsibility (Matthias 2004), and generate emergent ethical implications that may only become apparent long after implementation (Ananny and Crawford 2018). Moreover, their inherent bias toward quantifiable metrics systematically disadvantages qualitative human values that resist numerical reduction (Kearns and Roth 2019).

Existing frameworks, while valuable, fail to address these challenges comprehensively. Value-sensitive design approaches (Friedman, Hendry and Borning 2019) often lack mechanisms for handling dynamic value evolution. Stakeholder theories (Freeman 2010) struggle to account for the distributed nature of agency in complex systems. Ethical guidelines for AI and optimization systems (IEEE Global Initiative 2019; Floridi and Cowls 2018) typically treat ethics as a constraint rather than an integral aspect of system function. These limitations call for a new theoretical framework that can bridge the gap between philosophical ethics and technical implementation.

2 Philosophical Foundations

2.1 Value Pluralism in Technical Systems

Value pluralism in technical systems is not merely a practical challenge but a fundamental ontological condition (Berlin 1969). This builds on Berlin’s seminal work on value incommensurability, extending his insights from political philosophy to technological systems. Traditional optimization approaches attempt to reduce all values to quantifiable metrics, treating efficiency as the ultimate measure of system success. This reductionist approach fails because it misunderstands the nature of values in socio-technical systems (Taylor 1989; MacIntyre 1981). As Taylor argues in “Sources of the Self,” values are constitutive

of human identity and cannot be reduced to mere preferences or utilities.

Technical systems actively embody and promote values through their design and operation (Winner 1980). This embodiment occurs through algorithmic decision criteria, interface design choices, system architecture decisions, and data selection methods (Selbst et al. 2019). Each technical choice privileges certain values over others, creating a value landscape that shapes system behavior and stakeholder interaction. This embodiment is not neutral but actively influences how values are expressed and prioritized within the socio-technical system (Latour 2005).

2.2 Distributed Moral Agency

Agency in technical systems emerges from networks of interaction between human and technical components (Latour 2005). Drawing on Giddens' (Giddens 1984) structuration theory, system operators, users, affected stakeholders, and technical components all participate in complex networks of agency where responsibility and influence are distributed across multiple nodes (Floridi and Cowls 2018). This distribution creates patterns of interaction that cannot be reduced to simple chains of causation or responsibility (Johnson 2005).

Technical systems mediate agency through their design and operation (Verbeek 2011). Interface choices, automation levels, and feedback mechanisms all shape how human agency is expressed and constrained within the system. This mediation creates new forms of agency that emerge from the interaction between human and technical components, requiring new frameworks for understanding responsibility and ethical behavior (mittelstadt2016).

2.3 Dynamic Ethics

Ethics in technical systems operates dynamically, evolving through system operation and stakeholder interaction (Dewey 1922). This dynamic nature emerges from the continuous learning processes that occur as systems and stakeholders interact, adapt, and respond to changing conditions. Ethical implications evolve as systems learn and adapt, creating

new challenges and opportunities for ethical implementation (Floridi and Cowls 2018).

Ethical implementation requires sensitivity to context at multiple levels (Nissenbaum 2010). Local value systems, organizational constraints, temporal factors, and stakeholder relationships all influence how ethical principles manifest in practice. This context sensitivity demands continuous evaluation and adaptation of ethical frameworks as systems evolve and contexts change (Friedman, Hendry and Borning 2019).

The framework emphasizes continuous ethical evaluation through structured processes of value alignment, impact assessment, and adaptation (Kearns and Roth 2019). This evaluation must account for both immediate and long-term effects, direct and indirect impacts, and intended and emergent consequences of system operation (Yeung 2018).

3 Theoretical Framework

The Responsible Optimization Implementation Framework consists of three integrated components: the Value Integration Framework (VIF), the Stakeholder Agency Model (SAM), and the Implementation Dynamics Framework (IDF). This tripartite structure builds on established approaches to value-sensitive design (Friedman, Kahn Jr and Borning 2008; Hoven, Vermaas and Poel 2013) while extending them to address the specific challenges of optimization systems. Each component addresses a distinct aspect of ethical system implementation while maintaining coherent theoretical connections with the others (Dignum 2019).

3.1 Value Integration Framework

The Value Integration Framework provides systematic mechanisms for managing value pluralism in technical systems (Poel 2013), moving beyond the traditional treatment of values as abstract principles to operationalize them through distinct but interconnected spheres of consideration. This operationalization creates a dynamic structure where values interact and evolve while maintaining their essential characteristics (**manders2011**).

At the foundation of this framework lie technical values, which form the operational

core of system functionality (Introna 2007). Efficiency, reliability, performance, and scalability constitute not absolute metrics but rather relative measures whose significance emerges through their relationships with other value spheres (Simon 2015). This represents a fundamental departure from conventional approaches that treat technical values as primary and other values as constraints. Instead, technical values operate within a broader ethical context that shapes their expression and importance (Orlikowski 2007).

Social values provide the ethical infrastructure within which technical systems operate, establishing the boundaries and guidelines for system behavior (Barocas and Selbst 2016). Fairness, transparency, accountability, and broader societal impact require both quantitative and qualitative evaluation mechanisms. For instance, fairness demands not only statistical measures of outcome distribution but also qualitative assessment of procedural justice (Pasquale 2015). Similarly, transparency encompasses both technical explainability and stakeholder comprehension, recognizing that meaningful transparency requires both technical capability and human understanding.

3.2 Stakeholder Agency Model

The Stakeholder Agency Model addresses the distributed nature of moral agency in technical systems through a structured analysis of agency levels, interaction patterns, and power dynamics (Mitchell, Agle and Wood 1997). The model moves beyond traditional stakeholder theory (Donaldson and Preston 1995) by explicitly incorporating technical components as active mediators of agency (Latour 2005; Law 1992).

Systemic agency operates at institutional and societal levels through regulatory bodies, standards organizations, and professional associations (Yeung 2018). These entities shape the context within which technical systems operate, establishing boundaries for acceptable behavior and frameworks for evaluation. Their agency manifests through formal mechanisms like regulations and informal influences like professional norms (Dignum 2019).

Temporal agency acknowledges future stakeholders whose interests must be considered in current decision-making (Hoven, Vermaas and Poel 2013). This includes not only future users but also those who will inherit the consequences of current system operations.

Temporal agency requires mechanisms for representing and protecting future interests in present decisions (Nissenbaum 2010).

3.3 Implementation Dynamics Framework

The Implementation Dynamics Framework provides structured mechanisms for managing system evolution while maintaining ethical alignment (March 1991). This framework acknowledges the inherent tension between stability and adaptation in technical systems, providing sophisticated approaches to managing change while preserving core values (Argyris and Schön 1978).

Control structures exhibit common characteristics across successful implementations while adapting to specific operational contexts (desanctis1994). Multi-level monitoring spans technical and ethical oversight, providing comprehensive system observation (Yeung 2018). Integrated feedback mechanisms capture input from multiple stakeholder channels, enabling responsive system evolution (Friedman, Kahn Jr and Borning 2008). Adaptive governance structures support system evolution while maintaining ethical boundaries (Dignum 2019).

The framework implements these control structures through multiple complementary mechanisms. Oversight mechanisms monitor system behavior across technical, operational, and ethical dimensions, providing early warning of potential issues. Intervention protocols establish clear procedures for addressing misalignment when detected, ensuring that corrective actions maintain system stability while restoring ethical alignment. Evaluation systems assess both immediate outcomes and longer-term implications of system operation, while adaptation frameworks provide structured approaches to implementing necessary changes.

This comprehensive approach to implementation dynamics enables systems to evolve while maintaining ethical alignment, addressing one of the fundamental challenges in ethical system implementation. By providing structured mechanisms for managing system evolution, the framework supports both stability and adaptation, ensuring that technical advancement occurs within an ethical framework that preserves and promotes human

values.

4 Integration of Philosophy and Practice

The practical implementation of ROIF requires systematic integration of philosophical principles with operational requirements (**rouse2002**). This integration occurs across epistemological, ethical, and practical dimensions, each requiring specific mechanisms for translating theoretical insights into implementable procedures (Wimsatt 2007). The framework’s approach to integration acknowledges the inherent complexity of socio-technical systems while providing structured methods for managing this complexity in practice (**pickering1995**).

4.1 Epistemological Integration

Epistemological integration addresses the fundamental challenge of knowledge in complex socio-technical systems (Jasanoff 2004). The framework acknowledges inherent limitations in system knowledge while providing structured approaches to managing uncertainty (**giere2006**). This integration operates through several key mechanisms.

5 Case Studies

5.1 Clinical Decision Support Systems

Clinical decision support systems exemplify the core tension between technical optimization and professional autonomy (Bates et al. 2003). These systems demonstrate three key theoretical principles from ROIF:

First, they illustrate value pluralism through the simultaneous operation of technical values (diagnostic accuracy, processing speed), professional values (clinical expertise, professional judgment), and personal values (patient autonomy, dignity) (Ash, Berg and Coiera 2004). The Mayo Clinic’s implementation specifically demonstrates how these values can be integrated rather than traded off against each other (Friedman, Hendry and

Borning 2019).

Second, they manifest distributed agency through the creation of human-system decision networks (Verbeek 2011). The system’s confidence metrics and override protocols implement ROIF’s spectrum-based agency model, where both human and technical components contribute to decision-making while maintaining clear lines of responsibility (Johnson 2005).

5.2 Financial Trading Systems

The Implementation Dynamics Framework appears through sophisticated monitoring and adjustment mechanisms that maintain system alignment with ethical principles while enabling market efficiency (Pasquale 2015). Continuous monitoring systems track both technical performance and ethical alignment, while adjustment mechanisms enable rapid response to changing conditions. Stakeholder feedback integration occurs through multiple channels, from formal consultation processes to real-time market response analysis. System behavior evolution follows carefully designed pathways that maintain ethical boundaries while enabling necessary adaptation to changing market conditions (Dignum 2019).

5.3 Supply Chain Optimization Systems

Control structures exhibit common characteristics across successful implementations while adapting to specific operational contexts (desanctis1994). Multi-level monitoring spans technical and ethical oversight, providing comprehensive system observation (Yeung 2018). Integrated feedback mechanisms capture input from multiple stakeholder channels, enabling responsive system evolution (Friedman, Kahn Jr and Borning 2008). Adaptive governance structures support system evolution while maintaining ethical boundaries (Dignum 2019).

6 Cross-Case Analysis

A systematic analysis of these cases reveals fundamental patterns in how ROIF’s theoretical principles manifest across different implementation contexts (Eisenhardt 1989; Yin 2018). These patterns not only validate the framework’s theoretical foundations but also illuminate how abstract principles translate into concrete operational mechanisms (Miles, Huberman and Saldaña 2014).

6.1 Value Integration Manifestations

The cases demonstrate three distinct patterns of value integration that align with ROIF’s theoretical predictions (Stake 2005):

6.1.1 Hierarchical Value Structures

Healthcare implementations reveal how value hierarchies emerge not through predetermined rankings but through dynamic interaction between operational contexts and stakeholder needs (Sittig and Singh 2010). The Mayo Clinic’s system demonstrates how professional values (clinical judgment) and technical values (diagnostic accuracy) form complementary rather than competitive hierarchies (Bates et al. 2003). This aligns with ROIF’s prediction that value relationships emerge through system operation rather than abstract prioritization (Poel 2013).

6.1.2 Value Network Effects

Financial trading systems reveal how values operate in interconnected networks rather than linear relationships (MacKenzie 2006). Market efficiency, for instance, emerges as a product of multiple interacting values: transparency enables fair participation, which enhances market stability, which in turn supports efficiency (Pasquale 2015). This validates ROIF’s theoretical model of values as interconnected networks rather than isolated principles (Friedman, Hendry and Borning 2019).

6.1.3 Temporal Value Dynamics

Supply chain implementations demonstrate how value relationships evolve over time, supporting ROIF’s dynamic ethics principle (Seuring and Müller 2008). Initial efficiency-focused implementations evolve to incorporate broader value considerations as system impacts become apparent (Gold, Seuring and Beske 2010), showing how value understanding develops through system operation (Pagell and Wu 2009).

6.2 Agency Distribution Patterns

Agency network topologies exhibit domain-specific patterns that nonetheless conform to ROIF’s theoretical predictions (Latour 2005). Healthcare implementations typically develop hierarchical networks that reflect established professional authority structures while incorporating new forms of technical agency (Johnson 2005). Financial systems generate distributed networks where market participants interact through multiple channels, creating complex webs of influence and responsibility (MacKenzie 2006). Supply chain systems often evolve hybrid networks that combine hierarchical oversight with distributed operational authority, enabling both local autonomy and system-wide coordination (Lee 2004).

Agency preservation mechanisms show remarkable consistency across domains while adapting to specific operational requirements (Mitchell, Agle and Wood 1997). Explicit override mechanisms appear universally but manifest differently according to domain needs - from immediate clinical interventions in healthcare to circuit breakers in financial trading (Citron and Pasquale 2014). Graduated intervention levels provide nuanced responses to different situations, while feedback integration systems operate across varying temporal and spatial scales to maintain effective human oversight (Suchman 2007).

6.3 Implementation Dynamics

The evolution of ethical implementation across these cases reveals systematic patterns that validate and extend ROIF’s theoretical framework (Miles, Huberman and Saldaña 2014). These patterns manifest through learning trajectories, adaptation mechanisms,

and control structures, each demonstrating how ethical implementation develops through system operation (**nonaka1995**).

Control structures exhibit common characteristics across successful implementations while adapting to specific operational contexts (**desanctis1994**). Multi-level monitoring spans technical and ethical oversight, providing comprehensive system observation (Yeung 2018). Integrated feedback mechanisms capture input from multiple stakeholder channels, enabling responsive system evolution (Friedman, Kahn Jr and Borning 2008). Adaptive governance structures support system evolution while maintaining ethical boundaries (Dignum 2019).

6.4 Theoretical Implications

Implementation theory develops through analysis of how ethical systems evolve in practice (Orlikowski 2007). The evidence reveals predictable patterns in ethical implementation evolution, suggesting underlying principles that guide system development (March 1991). Learning occurs simultaneously across multiple dimensions, requiring integrated approaches to system development (**nonaka1995**). The necessity of matching control structure complexity to system complexity emerges as a fundamental principle (Senge 1990).

6.5 Framework Validation

The case studies provide compelling empirical validation of ROIF’s core theoretical claims, demonstrating how abstract principles manifest in practical implementation (Eisenhardt 1989; Yin 2018). This validation emerges across three fundamental dimensions: value pluralism, distributed agency, and dynamic ethics, each supporting the framework’s theoretical foundations while revealing nuances in practical application (Miles, Huberman and Saldaña 2014).

The distributed agency model receives particular validation through observed implementation patterns (Latour 2005). These patterns confirm ROIF’s conceptualization of agency as distributed across networks rather than residing in individual actors or com-

ponents (Johnson 2005). The evidence reveals how agency emerges through complex interactions between human and technical elements, creating sophisticated networks of influence and responsibility (Verbeek 2011). This validation supports the framework’s departure from traditional binary approaches to agency, demonstrating the practical utility of a more nuanced, network-based understanding (Matthias 2004).

7 Discussion and Implications

In the philosophy of technology, ROIF fundamentally reconceptualizes the relationship between values and technical systems (Verbeek 2011). The framework moves beyond traditional views of technology as value-neutral tools or mere carriers of embedded values, demonstrating instead how values actively shape and are shaped by technical system operation (Winner 1980). This dynamic, bi-directional relationship between values and technology emerges through the framework’s sophisticated treatment of value embodiment, where technical choices not only reflect but actively influence value expression and evolution (Latour 2005). This reconceptualization provides new theoretical tools for understanding how values manifest in increasingly complex technical environments (Poel 2013).

The practical implications of ROIF manifest across multiple dimensions of organizational and technical implementation, fundamentally reshaping how organizations approach ethical system deployment (Dignum 2019). At the system implementation level, ROIF transforms traditional development approaches by providing sophisticated mechanisms for embedding ethical considerations throughout the system lifecycle (Friedman, Kahn Jr and Borning 2008). Rather than treating ethical considerations as post-hoc constraints, these mechanisms enable organizations to integrate value sensitivity from the earliest stages of system design through to ongoing operation and evolution (Hoven, Vermaas and Poel 2013).

8 Conclusion

As technical systems continue to evolve in complexity and capability, the need for frameworks that can manage ethical implementation while supporting technical advancement becomes increasingly critical (Winner 1980). ROIF provides a robust foundation for meeting this challenge, offering both theoretical insight and practical guidance for organizations committed to responsible system implementation (Orlikowski 2007). Its emphasis on continuous learning and adaptation ensures its relevance in an evolving technological landscape (Argyris and Schön 1978), while its grounding in fundamental ethical principles provides stability amid rapid change (MacIntyre 1981).

The future of ethical system implementation will require continued refinement of frameworks like ROIF through both theoretical development and practical application (Suchman 2007). However, the core principles established here - the recognition of value pluralism (Berlin 1969), the importance of distributed agency (Latour 2005), and the necessity of dynamic ethics (Dewey 1922) - will remain essential guides for responsible system implementation. As we move forward, ROIF offers not just a framework for current implementation but a vision for how technical systems can evolve while maintaining and enhancing their commitment to human values and ethical principles (Friedman, Hendry and Borning 2019).

9 Future Research Directions

The development of ROIF opens critical pathways for future research that span both theoretical advancement and practical application (Miles, Huberman and Saldaña 2014). These research directions emerge from careful analysis of current framework implementation experiences and identification of areas requiring deeper understanding or more sophisticated approaches (Eisenhardt 1989).

9.1 Theoretical Development

ROIF’s theoretical innovations extend beyond traditional approaches to ethical system design (Verbeek 2011). The framework’s treatment of value pluralism as an operational reality rather than a philosophical challenge provides new ways to understand and manage value relationships in technical systems (Berlin 1969). Its spectrum-based approach to agency offers a more nuanced and practical model for human-system interaction than traditional binary frameworks (Johnson 2005). The dynamic ethics principle, supported by structured adaptation mechanisms, enables systems to evolve while maintaining ethical alignment (March 1991).

9.2 Practical Applications

The practical implications of ROIF manifest across multiple dimensions of organizational and technical implementation, fundamentally reshaping how organizations approach ethical system deployment (Dignum 2019). At the system implementation level, ROIF transforms traditional development approaches by providing sophisticated mechanisms for embedding ethical considerations throughout the system lifecycle (Friedman, Kahn Jr and Borning 2008). Rather than treating ethical considerations as post-hoc constraints, these mechanisms enable organizations to integrate value sensitivity from the earliest stages of system design through to ongoing operation and evolution (Hoven, Vermaas and Poel 2013).

Organizational governance receives substantial practical enhancement through ROIF’s structured approaches to ethical oversight (Mitchell, Agle and Wood 1997). The framework provides detailed mechanisms for stakeholder engagement that enable organizations to systematically incorporate diverse perspectives while maintaining operational coherence (Freeman 2010). These mechanisms extend beyond traditional consultation models to create dynamic feedback systems that continuously inform system evolution (**wenger1998**). Value alignment processes operate across organizational boundaries, ensuring consistent ethical implementation even in complex, distributed systems (**nonaka1995**).

References

- Ananny, Mike and Kate Crawford (2018). ‘Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability’. In: *New Media & Society* 20.3, pp. 973–989.
- Argyris, Chris and Donald A. Schön (1978). *Organizational Learning: A Theory of Action Perspective*. Addison-Wesley.
- Ash, Joan S., Marc Berg and Enrico Coiera (2004). ‘Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-related Errors’. In: *Journal of the American Medical Informatics Association* 11.2, pp. 104–112.
- Barocas, Solon and Andrew D. Selbst (2016). ‘Big Data’s Disparate Impact’. In: *California Law Review* 104, pp. 671–732.
- Bates, David W. et al. (2003). ‘Ten Commandments for Effective Clinical Decision Support: Making the Practice of Evidence-based Medicine a Reality’. In: *Journal of the American Medical Informatics Association* 10.6, pp. 523–530.
- Berlin, Isaiah (1969). *Four Essays on Liberty*. Oxford University Press.
- Citron, Danielle Keats and Frank Pasquale (2014). ‘The Scored Society: Due Process for Automated Predictions’. In: *Washington Law Review* 89, pp. 1–33.
- Dewey, John (1922). *Human Nature and Conduct: An Introduction to Social Psychology*. Henry Holt and Company.
- Dignum, Virginia (2019). ‘Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way’. In: *Springer Nature*.
- Donaldson, Thomas and Lee E. Preston (1995). ‘The Stakeholder Theory of the Corporation: Concepts, Evidence, and Implications’. In: *Academy of Management Review* 20.1, pp. 65–91.
- Eisenhardt, Kathleen M. (1989). ‘Building Theories from Case Study Research’. In: *Academy of Management Review* 14.4, pp. 532–550.
- Floridi, Luciano and Josh Cowls (2018). ‘A Unified Framework of Five Principles for AI in Society’. In: *Harvard Data Science Review* 1.1.

- Freeman, R. Edward (2010). *Strategic Management: A Stakeholder Approach*. Cambridge University Press.
- Friedman, Batya, David G. Hendry and Alan Borning (2019). ‘A Survey of Value Sensitive Design Methods’. In: *Foundations and Trends in Human-Computer Interaction* 11.2, pp. 63–125.
- Friedman, Batya, Peter H. Kahn Jr and Alan Borning (2008). ‘Value Sensitive Design and Information Systems’. In: *The Handbook of Information and Computer Ethics*, pp. 69–101.
- Giddens, Anthony (1984). *The Constitution of Society: Outline of the Theory of Structuration*. University of California Press.
- Gold, Stefan, Stefan Seuring and Philip Beske (2010). ‘Sustainable Supply Chain Management and Inter-Organizational Resources: A Literature Review’. In: *Corporate Social Responsibility and Environmental Management* 17.4, pp. 230–245.
- Hoven, Jeroen van den, Pieter E. Vermaas and Ibo van de Poel (2013). ‘Handbook of Ethics, Values, and Technological Design’. In: *Springer*, pp. 1–875.
- IEEE Global Initiative (2019). ‘Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems’. In: *IEEE Standards Association* 1.
- Introna, Lucas D. (2007). ‘Maintaining the reversibility of foldings: Making the ethics (politics) of information technology visible’. In: *Ethics and Information Technology* 9.1, pp. 11–25.
- Jasanoff, Sheila (2004). *States of Knowledge: The Co-Production of Science and Social Order*. Routledge.
- Johnson, Deborah G. (2005). ‘Computer systems: Moral entities but not moral agents’. In: *Ethics and Information Technology* 8.4, pp. 195–204.
- Kearns, Michael and Aaron Roth (2019). *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press.
- Latour, Bruno (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.

- Law, John (1992). 'Notes on the Theory of the Actor-Network: Ordering, Strategy, and Heterogeneity'. In: *Systems Practice* 5.4, pp. 379–393.
- Lee, Hau L. (2004). 'The Triple-A Supply Chain'. In: *Harvard Business Review* 82.10, pp. 102–112.
- MacIntyre, Alasdair (1981). *After Virtue: A Study in Moral Theory*. University of Notre Dame Press.
- MacKenzie, Donald (2006). 'An Engine, Not a Camera: How Financial Models Shape Markets'. In: *MIT Press*.
- March, James G. (1991). 'Exploration and Exploitation in Organizational Learning'. In: *Organization Science* 2.1, pp. 71–87.
- Matthias, Andreas (2004). 'The responsibility gap: Ascribing responsibility for the actions of learning automata'. In: *Ethics and Information Technology* 6.3, pp. 175–183.
- Miles, Matthew B., A. Michael Huberman and Johnny Saldaña (2014). *Qualitative Data Analysis: A Methods Sourcebook*. Sage Publications.
- Mitchell, Ronald K., Bradley R. Agle and Donna J. Wood (1997). 'Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts'. In: *Academy of Management Review* 22.4, pp. 853–886.
- Nissenbaum, Helen (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press.
- Orlikowski, Wanda J. (2007). 'Sociomaterial Practices: Exploring Technology at Work'. In: *Organization Studies* 28.9, pp. 1435–1448.
- Pagell, Mark and Zhaohui Wu (2009). 'Building a More Complete Theory of Sustainable Supply Chain Management Using Case Studies of 10 Exemplars'. In: *Journal of Supply Chain Management* 45.2, pp. 37–56.
- Pasquale, Frank (2015). 'The Black Box Society: The Secret Algorithms That Control Money and Information'. In: *Harvard University Press*.
- Poel, Ibo van de (2013). 'Translating Values into Design Requirements'. In: *Philosophy and Engineering: Reflections on Practice, Principles and Process*, pp. 253–266.

- Selbst, Andrew D. et al. (2019). *Fairness and Abstraction in Sociotechnical Systems*. ACM, pp. 59–68.
- Senge, Peter M. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. Doubleday.
- Seuring, Stefan and Martin Müller (2008). ‘From a Literature Review to a Conceptual Framework for Sustainable Supply Chain Management’. In: *Journal of Cleaner Production* 16.15, pp. 1699–1710.
- Simon, Judith (2015). ‘Value-Sensitive Design and Responsible Research and Innovation’. In: *The Ethics of Technology: Methods and Approaches*, pp. 219–235.
- Sittig, Dean F. and Hardeep Singh (2010). ‘A New Sociotechnical Model for Studying Health Information Technology in Complex Adaptive Healthcare Systems’. In: *Quality and Safety in Health Care* 19.Suppl 3, pp. i68–i74.
- Stake, Robert E. (2005). ‘Qualitative Case Studies’. In: *The Sage Handbook of Qualitative Research*, pp. 443–466.
- Suchman, Lucy (2007). ‘Human-Machine Reconfigurations: Plans and Situated Actions’. In: *Cambridge University Press*.
- Taylor, Charles (1989). *Sources of the Self: The Making of the Modern Identity*. Harvard University Press.
- Verbeek, Peter-Paul (2011). ‘Moralizing Technology: Understanding and Designing the Morality of Things’. In: *University of Chicago Press*.
- Wimsatt, William C. (2007). *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard University Press.
- Winner, Langdon (1980). *Do Artifacts Have Politics?* Vol. 109. 1. Daedalus, pp. 121–136.
- Yeung, Karen (2018). ‘Algorithmic regulation: A critical interrogation’. In: *Regulation & Governance* 12.4, pp. 505–523.
- Yin, Robert K. (2018). *Case Study Research and Applications: Design and Methods*. Sage Publications.

Zarsky, Tal Z. (2016). ‘The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making’. In: *Science, Technology, & Human Values* 41.1, pp. 118–132.