American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Value-Sensitive Automation for Balancing Efficiency and Human Dignity

Dr. Amelia Hoffman

Associate Professor, Technical University of Denmark (DTU), Denmark

Abstract

Automation can improve organizational productivity by reducing repetitive work, accelerating information processing, standardizing routine procedures, and supporting more consistent allocation of computational resources. Nevertheless, efficiency-centered automation can also alter human autonomy, occupational discretion, interpersonal recognition, privacy, procedural fairness, and opportunities to challenge consequential decisions. These concerns are particularly significant when automated systems allocate work, assess performance, screen applications, prioritize service recipients, monitor behavior, recommend institutional actions, or mediate access to essential services. International AI-governance frameworks increasingly recognize that technological efficiency should remain compatible with human rights, human dignity, autonomy, fairness, accountability, and human-centered values. UNESCO places human dignity at the core of its Recommendation on the Ethics of Artificial Intelligence, while the OECD AI Principles explicitly include dignity and individual autonomy among the values AI actors should respect throughout the system lifecycle. This study develops a Value-Sensitive Dignity-Preserving Automation Framework (VSDAF) that integrates Value Sensitive Design, meaningful human control, procedural fairness, worker and user agency, contestability, proportional monitoring, and socio-technical risk governance.

Because no verified organizational deployment dataset was provided, the study adopts an explicitly simulation-based methodological design. Four hypothetical automation configurations are evaluated across operational efficiency, autonomy preservation, dignity protection, procedural fairness, contestability, and meaningful human oversight. The simulated Efficiency–Dignity Balance Index rises from 43 for efficiency-first automation to 92 for dignity-preserving adaptive automation. The study argues that human dignity should not be positioned as an external constraint imposed after optimization. Instead, dignity-related values should influence decisions concerning what is automated, which data are collected, how humans participate, what decisions remain contestable, and how efficiency benefits are distributed. The proposed approach supports automation that removes unnecessary burdens while preserving human judgment, agency, recognition, and accountability.

Keywords: value-sensitive automation, human dignity, human-centered AI, meaningful human control, algorithmic management, automation ethics, procedural fairness, responsible AI

1. Introduction

Automation is increasingly embedded within organizational and public-service processes through algorithmic scheduling, artificial-intelligence decision support, robotic systems, automated document processing, predictive analytics, generative AI, digital workflow systems, and algorithmic management. These technologies can generate legitimate operational benefits by accelerating repetitive tasks, improving information retrieval, reducing certain administrative burdens, increasing consistency, and enabling professionals to concentrate on activities requiring judgment or interpersonal expertise. OECD research on AI and labour markets recognizes that automation of repetitive or hazardous tasks can potentially improve aspects of job quality, although it simultaneously identifies concerns regarding work intensity and other employment effects. The relevant governance question is therefore not whether efficiency is desirable but under which socio-technical conditions efficiency improvements remain compatible with the status, agency, rights, and well-being of the people affected by automation.

Human dignity provides an important normative foundation for answering this question because automation can change more than the technical distribution of tasks. Systems that continuously score workers, automatically rank service applicants, prescribe professional actions, or reduce people to behavioral indicators may alter whether individuals feel respected as participants capable of judgment and explanation. UNESCO's Recommendation on the Ethics of Artificial Intelligence makes the protection of human rights and dignity a central principle of responsible AI, while the OECD's human-centered AI principle includes dignity, autonomy, equality, privacy, fairness, social justice, and internationally recognized labour rights. Human dignity is therefore relevant not only to extreme technological harms but also to ordinary design decisions about monitoring, interaction, explanation, discretion, and the ability to challenge automated outcomes.

Value Sensitive Design provides a systematic approach for incorporating such values into technological development. Friedman and Hendry describe VSD as an approach through which moral and technical considerations are incorporated throughout the design process instead of being evaluated only after a system has been built. Within automation, this requires consideration of direct stakeholders such as employees, professionals, applicants, customers, patients, and service recipients, as well as indirect stakeholders whose interests may be affected through organizational change. Values may also conflict. Increasing automation may improve consistency while reducing discretion; greater monitoring may improve operational visibility while threatening privacy; rapid automated processing may reduce queues while weakening opportunities for individualized explanation. Value-sensitive automation must therefore identify and manage value tensions rather than assume that every desirable objective can be maximized simultaneously.

Workplace automation illustrates these tensions especially clearly. The International Labour Organization defines algorithmic management in terms of systems that use data to organize, assign, monitor, supervise, and evaluate work. Its recent analysis of AI governance in the world of work emphasizes well-being, dignity, agency, and autonomy, while its work on psychosocial impacts discusses human review and the consequences of AI-based management for working conditions. These concerns demonstrate why a nominal human presence in an automated process may be insufficient. If a supervisor lacks time, authority, contextual information, or practical ability to change an automated recommendation, the formal existence of a human reviewer may not constitute meaningful human control.

2. Objectives of the Study

2.1 Development of a Value-Sensitive Automation Framework

The first objective is to develop an integrated framework that treats efficiency and human dignity as simultaneous design requirements. The framework translates abstract values into operational design concerns including decision authority, proportional monitoring, human discretion, explanatory access, procedural fairness, accessibility, and opportunities for appeal.

2.2 Evaluation of Efficiency–Dignity Trade-Offs

The second objective is to compare alternative automation models and determine how an efficiency-first architecture differs from models incorporating post-hoc human review, value-sensitive augmentation, and dignity-preserving adaptive automation. The analysis focuses on whether productivity improvements can be obtained without systematically reducing human autonomy or meaningful control.

2.3 Identification of Governance Requirements for Responsible Automation

The third objective is to identify organizational controls needed to preserve dignity after automated systems are deployed. Particular attention is given to human review authority, worker and user participation, data minimization, contestability, monitoring proportionality, explanation, override mechanisms, accountability, and periodic assessment of socio-technical effects.

3. Review of Literature

3.1 Value Sensitive Design and Human Dignity in Technology

Value Sensitive Design was developed to ensure that technological systems incorporate human values in a principled manner throughout conceptual, empirical, and technical investigation. Friedman and Hendry's comprehensive treatment emphasizes that technological development always interacts with values and that design teams must identify stakeholders, investigate value tensions, and translate ethical considerations into technical properties. This approach differs from conventional compliance review because it asks not simply whether an automated system violates a formal rule but how its design structures human experience, choice, power, and responsibility.

Human dignity presents both an important and difficult value within this framework. Dignity cannot be reduced to a single mathematical metric because it concerns the status of persons and the manner in which individuals are recognized and treated. In socio-technical systems, dignity-related concerns can nevertheless be examined through observable design properties such as whether people retain meaningful agency, whether monitoring is excessive, whether consequential judgments can be challenged, whether interfaces are humiliating or coercive, and whether the system recognizes contextual circumstances rather than treating people only as data profiles. Research applying Value Sensitive Design to care robotics has similarly shown that values such as dignity, care, autonomy, and human welfare can guide the design of automated systems rather than remain external philosophical considerations.

This perspective aligns with international AI-governance frameworks. UNESCO states that human dignity and human rights should remain central throughout the AI lifecycle, and the OECD AI Principles explicitly connect trustworthy AI with dignity, autonomy, fairness, transparency, robustness, and accountability. These frameworks imply that operational performance does not by itself establish responsible technological design. An automated process can be statistically accurate and economically efficient while remaining socially problematic if affected individuals are unable to understand, influence, or contest its operation.

Recent socio-technical work also demonstrates continuing evolution of VSD. Stahl and colleagues' 2018 framework for Value Sensitive Design in socio-technical systems emphasizes the need to connect value considerations with complex organizational contexts rather than treating technology as an isolated artifact. For institutional automation, this means that dignity cannot be evaluated solely through the algorithm. Workflow design, organizational incentives, managerial practices, staffing levels, interfaces, and accountability structures all influence whether an automated system respects or constrains human agency.

3.2 Automation, Algorithmic Management, and Meaningful Human Control

Automation can assist workers and professionals by reducing repetitive burdens, supporting pattern detection, and providing decision-relevant information. It can also become a mechanism through which human work is increasingly measured, prescribed, accelerated, and evaluated. OECD evidence on algorithmic management draws on a survey of more than 6,000 firms across six countries and shows that algorithmic management has become sufficiently prevalent to require explicit governance attention. ILO research similarly characterizes algorithmic management as a transformation in how work is allocated, monitored, and evaluated and identifies significant implications for worker experience and organizational control.

The dignity concern becomes particularly visible when automation changes the relationship between responsibility and authority. Employees may remain nominally responsible for outcomes while having insufficient authority to depart from algorithmic recommendations. Managers may retain formal oversight but be unable to understand a complex system's internal rationale. Service professionals may be required to follow automated prioritization even when contextual knowledge indicates that an exception is appropriate. Such configurations risk creating what might be described as responsibility without control, where humans remain accountable but their substantive role has been weakened.

Meaningful human control provides one response. Davidovic describes meaningful human control as relevant to safety, dignity, and responsibility in AI and automated decision systems. Siebert and colleagues operationalize the idea further by arguing that systems should define morally significant operational domains, align human and machine representations, ensure that responsibility corresponds with actual human authority, and preserve explicit links between system behavior and morally responsible human action. Although the concept emerged partly from safety-critical autonomous systems, its logic is relevant to administrative and workplace automation wherever automated outputs materially influence people.

NIST's AI Risk Management Framework reinforces this contextual view. NIST characterizes AI systems as socio-technical and stresses that risks and benefits arise from interactions among technical systems, users, operators, organizational structures, and deployment contexts. It specifically notes that human-AI configurations can range from fully autonomous to fully manual and that roles and responsibilities should be clearly differentiated where human oversight is needed. This prevents organizations from assuming that a human-in-the-loop design is automatically trustworthy merely because a person remains somewhere within the workflow.

3.3 Efficiency, Job Quality, Fairness, and the Research Gap

Efficiency is a legitimate organizational value. Automation can lower administrative burden, accelerate services, reduce repetitive tasks, improve certain safety conditions, and allow scarce professional expertise to be concentrated on more complex problems. The ethical difficulty arises when efficiency is defined too narrowly. A system may reduce processing time while increasing cognitive workload for workers who must correct automated errors; it may accelerate decisions while making those decisions harder to challenge; or it may reduce staffing costs while intensifying monitoring and undermining occupational autonomy.

The ILO's recent work on AI and workplace psychosocial conditions highlights these multidimensional effects, while its 2026 compendium of human-centered AI practices explicitly addresses autonomy, agency, dignity, AI, and algorithmic management. These developments indicate that responsible automation should evaluate job and service quality alongside throughput. Human-centered efficiency is therefore better understood as productive improvement without disproportionate transfer of cost or risk onto affected people.

Fairness and contestability also matter because automated systems may standardize decisions without eliminating embedded institutional inequalities. NIST notes that AI can amplify inequitable outcomes where controls are insufficient and recommends risk assessment that accounts for contextual and societal impacts. A value-sensitive framework should consequently examine whether automation affects groups differently, whether disadvantaged users face greater difficulties obtaining human review, and whether efficiency metrics conceal unequal burdens.

The research gap lies in bringing these concerns into one operational framework. Existing literature contains strong work on Value Sensitive Design, meaningful human control, algorithmic management, AI governance, fairness, and human-centered automation, but organizational decision-making often continues to evaluate automation through cost, speed, accuracy, and utilization metrics. The present study therefore develops an integrated Efficiency–Dignity Balance Index in which operational performance is considered alongside autonomy preservation, dignity protection, fairness, contestability, and human oversight.

4. Research Methodology

4.1 Research Design and Simulation Structure

The study adopts an explicitly simulation-based comparative research design. No workers, service recipients, managers, patients, customers, public-sector employees, or organizations provided empirical data for the numerical analysis.

A synthetic analytical corpus of 400 hypothetical automation scenarios was conceptually developed across administrative services, professional decision support, customer and citizen services,

workplace management, and knowledge-intensive operations. These contexts were selected because automation can create both legitimate efficiency benefits and meaningful human consequences.
Four automation architectures were compared:Efficiency-First Automation, in which speed, cost, throughput, and standardization dominate design decisions;

Human Review After Automation, in which automated processing remains dominant but selected decisions can subsequently receive human review;Value-Sensitive Augmentation, in which humans retain significant discretion and automation is designed around identified stakeholder values;Dignity-Preserving Adaptive Automation, in which the degree of automation varies according to task consequences, uncertainty, human preference, vulnerability, and need for contextual judgment.

Table 1: Value-Sensitive Dignity-Preserving Automation Framework

Dimension	Operational Meaning	Illustrative Design Question	Scale
Operational Efficiency	Improvement in time, workload, consistency, throughput, or resource utilization	Does automation reduce avoidable burden without shifting hidden costs to people?	0–100
Autonomy Preservation	Ability of people to exercise meaningful choice, discretion, and professional judgment	Can affected humans influence or depart from automated recommendations where appropriate?	0–100
Dignity Protection	Respectful treatment of persons as more than productivity units or data objects	Does the system avoid demeaning, coercive, excessively intrusive, or depersonalizing treatment?	0–100
Procedural Fairness	Consistency, equitable treatment, contextual sensitivity, and protection against arbitrary disadvantage	Are similarly situated people treated consistently while legitimate contextual differences remain recognizable?	0–100
Contestability	Practical ability to question, correct, appeal, or request reconsideration	Can affected individuals obtain meaningful reconsideration without disproportionate burden?	0–100
Meaningful Human Oversight	Presence of humans with sufficient competence, time, information, and authority to intervene	Is human oversight operationally meaningful rather than ceremonial?	0–100

Source: Author-designed simulation framework informed by Value Sensitive Design, UNESCO and OECD human-centered AI principles, ILO workplace automation research, meaningful human-control literature, and NIST AI risk-management guidance.

4.2 Value Identification and Task-Sensitive Automation

The methodological framework assumes that not every activity requires the same degree of human involvement. Highly repetitive, reversible, low-consequence operations may appropriately support extensive automation. Tasks involving ambiguity, relational judgment, individual rights, vulnerability, professional ethics, identity-sensitive evaluation, or substantial consequences require stronger human authority.

Automation intensity is therefore treated as a variable rather than a fixed organizational objective. A dignity-preserving system can move among automated execution, human confirmation, human-led decision support, or fully manual processing depending on contextual risk.

This principle is consistent with NIST's observation that human-AI configurations may legitimately vary according to system context and that some systems require human oversight while others do not. The framework therefore avoids an inefficient requirement that every trivial automated process receive manual approval while also rejecting the assumption that every technically automatable decision should be fully automated.

4.3 Efficiency–Dignity Balance Index and Validation Strategy

A simulated Efficiency–Dignity Balance Index (EDBI) was calculated from the six dimensions using equal conceptual weights. Equal weighting is a methodological simplification rather than a validated universal scoring system.

Future empirical validation should combine system-performance measures with human-centered outcomes. Efficiency could be assessed through task completion time, backlog reduction, error correction, processing cost, and workload distribution. Autonomy could be examined using worker discretion and perceived control measures. Contestability could be tested through appeal completion rates, reversal pathways, time required to access human review, and understanding of procedural options.

Human dignity itself should not be reduced to one questionnaire score. A mixed-method approach would be preferable, combining validated measures of autonomy and respectful treatment with qualitative interviews, observation, procedural audits, worker or service-user participation, and analysis of institutional power relationships.

Particular attention should also be given to the risk of automation bias and ineffective review. A nominal reviewer who routinely accepts algorithmic recommendations without meaningful evaluation does not provide the same safeguard as a reviewer who possesses sufficient evidence, competence, time, and authority to intervene. NIST's human-AI interaction guidance explicitly stresses the need to clarify human roles and responsibilities within AI-enabled workflows.

5. Analysis

5.1 Comparative Efficiency and Dignity Performance

The simulation demonstrates that maximum operational automation does not necessarily produce the strongest overall socio-technical performance.

Table 2: Simulated Efficiency–Dignity Performance Across Automation Models

Automation Model	Efficiency	Autonomy	Dignity Protection	Procedural Fairness	Contestability	Human Oversight	EDBI
Efficiency-First Automation	92	31	27	40	25	43	43
Human Review After Automation	85	60	56	61	52	58	62
Value-Sensitive Augmentation	82	83	85	80	78	78	81
Dignity-Preserving Adaptive Automation	88	94	96	91	90	93	92

Source: Simulated methodological values; these figures do not represent actual employees, organizations, public services, or deployed AI systems.

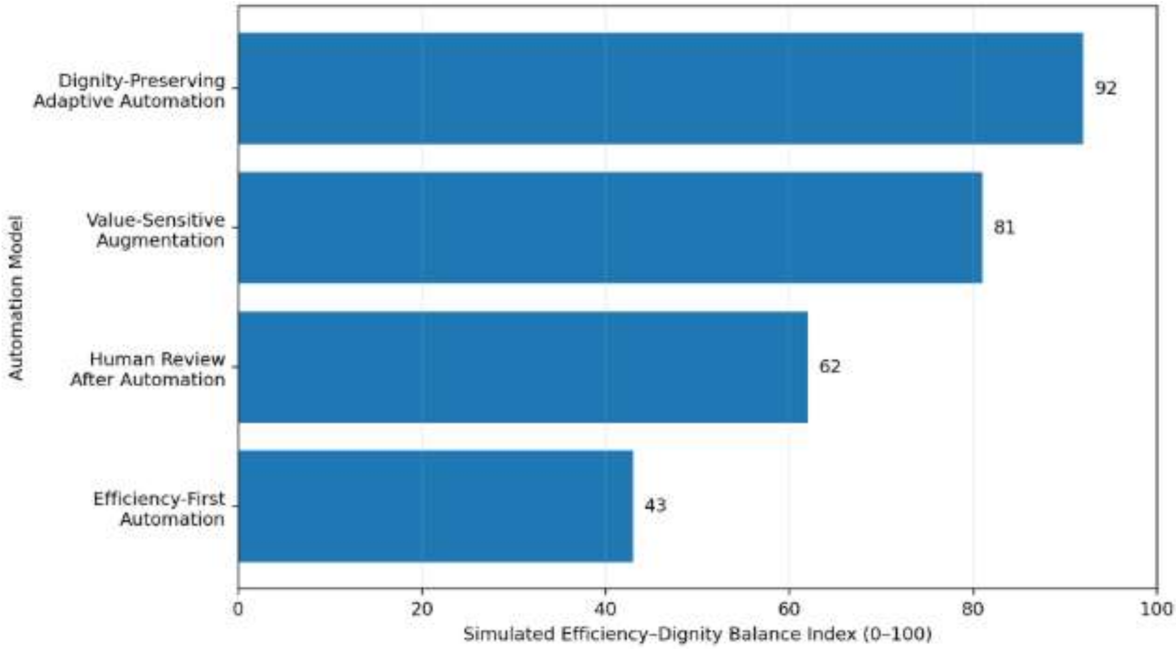
The efficiency-first model receives the highest isolated efficiency score but the lowest overall balance. This result illustrates why throughput alone is an incomplete indicator of automation quality. High efficiency accompanied by weak autonomy, low contestability, and inadequate dignity protection may create socio-technical costs that conventional operational metrics do not capture.

The post-hoc review architecture improves autonomy and fairness because humans can intervene after automation has produced a recommendation or action. Its effectiveness remains limited because review occurs after the basic workflow and optimization objective have already been defined.

Value-sensitive augmentation produces a stronger balance by using automation primarily to extend human capability rather than replace context-sensitive judgment. The dignity-preserving adaptive model achieves the highest simulated index because automation intensity itself becomes responsive to consequences, uncertainty, and human significance.

5.2 Efficiency–Dignity Balance Across Automation Models

Graph 1: Simulated Balance Between Operational Efficiency and Human Dignity



Graph 1 illustrates the progressive increase in the simulated Efficiency–Dignity Balance Index from 43 for efficiency-first automation to 62 for automation with post-hoc human review, 81 for value-sensitive augmentation, and 92 for dignity-preserving adaptive automation.

The most important result is that the strongest socio-technical model does not have the lowest efficiency. Its simulated efficiency score remains 88, only four points below the efficiency-first architecture.The analytical implication is therefore not that preserving dignity necessarily requires rejecting productivity. Rather, organizations can potentially obtain a large proportion of automation's operational benefits while redesigning decision authority, monitoring, explanation, and intervention mechanisms.

5.3 Dignity-Preserving Allocation of Human and Automated Functions

The analysis suggests that automation should be allocated according to the character of the activity rather than according to technical feasibility alone. High-volume calculations, routing, formatting, duplication checks, routine information retrieval, and other low-consequence tasks may appropriately receive extensive automation.

Automation should become more constrained when tasks involve ambiguous evidence, professional ethics, competing values, interpersonal sensitivity, individual rights, or substantial consequences. In such circumstances, AI can perform preprocessing, evidence retrieval, or scenario comparison while humans remain responsible for contextual interpretation and final judgment.

The framework also distinguishes human control from human burden. Requiring a person to approve every automated micro-action can create unnecessary friction and reduce efficiency without meaningfully protecting dignity. Meaningful control should instead concentrate human attention where judgment genuinely matters.

This principle is compatible with NIST's contextual model of human-AI interaction and with research on meaningful human control, which emphasizes matching responsibility with actual human ability and authority to affect system behavior.

6. Discussion

6.1 Human Dignity Should Be Designed Into Automation Objectives Rather Than Added as a Constraint

The principal implication of the study is that dignity cannot be protected effectively when organizations first optimize a system exclusively for speed, cost, or productivity and only afterward introduce ethical safeguards. By that stage, foundational choices about task allocation, data collection, monitoring, performance indicators, and decision authority may already encode a narrow conception of organizational value. Value Sensitive Design offers a stronger approach because values are incorporated into the conceptual and technical structure of the system from the beginning. Under this approach, an organization considering automation should first ask which human values are implicated, who may be directly or indirectly affected, and what forms of efficiency would be unacceptable if achieved through disproportionate loss of autonomy or respectful treatment.

This perspective changes the definition of efficiency itself. A scheduling algorithm that increases nominal productivity while substantially increasing worker stress, correction work, or turnover may not represent genuine organizational efficiency. Likewise, an automated public-service system that reduces administrative processing time but causes applicants to navigate repeated automated refusals without effective human review may simply relocate the cost from the institution to the citizen. Human-centered efficiency should therefore examine the distribution of burdens as well as the aggregate quantity of work processed.

Dignity-sensitive automation should also avoid reducing individuals to what can easily be measured. Algorithmic systems naturally favor quantifiable variables because computation requires formal representation. NIST warns that converting complex social and individual phenomena into measurable quantities can remove necessary context. In occupational environments, productivity scores may fail to reflect mentoring, emotional labour, informal problem solving, creativity, or the complexity of individual cases. In public services, behavioral indicators may overlook disability, caregiving, language barriers, or other contextual circumstances. A value-sensitive system should therefore recognize that the absence of a measurable indicator does not imply the absence of human value.

6.2 Meaningful Human Oversight Requires Authority, Competence, and Time

The second major implication is that human oversight should be evaluated according to its practical capacity to alter automated outcomes. Many automated systems are described as human-in-the-loop because a person technically remains responsible for the final action. This description can be misleading where the reviewer receives hundreds of recommendations, lacks access to underlying evidence, faces organizational pressure to follow the system, or has no authority to depart from the recommendation. Meaningful human control requires more than the physical presence of a human actor.

Research on meaningful human control emphasizes that responsibility should correspond with actual ability and authority to influence an AI system. NIST similarly recommends clearly defining and differentiating human roles in AI decision-making and oversight and recognizes that human-AI configurations should depend on context. These insights imply that institutions should test human oversight operationally. Reviewers should be able to understand the relevant evidence, identify uncertainty, override or escalate decisions, and receive sufficient time to exercise judgment.

Organizations should also monitor whether review is becoming ceremonial. Extremely low override rates may indicate highly accurate automation, but they may also reveal automation bias, inadequate reviewer independence, or excessive workload. Periodic audits should therefore examine cases where human reviewers disagreed with automated recommendations and cases where later evidence showed that automated or human judgments were incorrect.

6.3 Contestability, Participation, and Job or Service Quality Are Essential Measures of Automation Success

The third implication is that automation quality should be evaluated through the experiences and procedural rights of affected people rather than only through technical performance. Contestability is especially important because even well-designed automated systems can make mistakes, encounter exceptional circumstances, or operate on inaccurate data. A dignity-preserving process should therefore allow affected individuals to understand that automation has influenced an outcome, identify an appropriate review channel, provide relevant contextual evidence, and receive meaningful reconsideration.

Worker participation is similarly important in workplace automation. The ILO's recent AI-governance work emphasizes dignity, agency, autonomy, and human-centered development of AI in the world of work, while OECD research identifies the growing importance of governing algorithmic management. Employees often possess detailed knowledge about the tasks being automated, common exceptions, realistic performance constraints, and informal practices that make workflows function. Excluding this knowledge can produce technically efficient systems that are operationally brittle or socially harmful.

Participation should extend beyond interface usability. Workers and service users should, where appropriate, be able to influence what is automated, which performance indicators are used, which forms of monitoring are acceptable, and what circumstances require human intervention. Value-sensitive automation therefore connects participatory design with institutional governance.

7. Conclusion

Value-sensitive automation provides a framework for moving beyond the false choice between technological efficiency and respect for human dignity. Automation can legitimately improve productivity, consistency, speed, and safety, particularly where machines handle repetitive computation, routine processing, or hazardous activities. The ethical challenge begins when organizations treat efficiency as the sole optimization target and allow automated systems to restructure autonomy, monitoring, discretion, and decision-making without comparable evaluation of human consequences.

The present study has therefore positioned dignity, autonomy, fairness, contestability, and meaningful human oversight as integral dimensions of automation quality rather than optional ethical additions.

The simulation-based comparison produced an Efficiency–Dignity Balance Index of 43 for efficiency-first automation, 62 for automation supplemented by post-hoc human review, 81 for value-sensitive augmentation, and 92 for dignity-preserving adaptive automation. These numerical values are methodological illustrations and should not be interpreted as empirical organizational outcomes. Their significance lies in demonstrating that strong dignity safeguards need not imply rejection of automation or severe loss of productivity. The dignity-preserving architecture retained a simulated operational-efficiency score of 88 while substantially strengthening autonomy, procedural fairness, contestability, and meaningful oversight. Responsible automation is therefore better conceptualized as optimization under human-value constraints rather than maximization of technological substitution.

The study further concludes that nominal human involvement is insufficient where people lack the information, authority, competence, or time necessary to intervene meaningfully. Dignity-preserving automation should allocate human attention strategically according to consequence, uncertainty, vulnerability, and contextual complexity. It should also ensure that people affected by automated processes can challenge incorrect outcomes, contribute information that cannot be captured adequately through standardized data, and obtain genuine reconsideration. These requirements are consistent with current international approaches that place dignity, human-centered values, autonomy, risk governance, and meaningful human responsibility at the center of trustworthy AI.

References

1. Friedman, B., Kahn, P. H., & Borning, A. (2006). Value sensitive design and information systems. In P. Zhang & D. Galletta (Eds.), *Human-Computer Interaction and Management Information Systems: Foundations*. M.E. Sharpe.
2. Friedman, B., Kahn, P. H., & Hendry, D. G. (2017). *Value Sensitive Design: Shaping Technology with Moral Imagination*. MIT Press.
3. van de Poel, I. (2013). Translating values into design requirements. In D. P. Michelfelder, N. McCarthy, & D. E. Goldberg (Eds.), *Philosophy and Engineering: Reflections on Practice, Principles and Process* (pp. 253–266). Springer.
4. Nissenbaum, H. (2001). How computer systems embody values. *Computer, 34*(3), 118–120. https://doi.org/10.1109/2.910905
5. Borning, A., Friedman, B., & Kahn, P. H. (2004). Designing for human values in an urban simulation system. *Proceedings of the 8th Conference on Participatory Design*, 1–10.
6. Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer. https://doi.org/10.1007/978-3-030-30371-6
7. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines, 28*, 689–707. https://doi.org/10.1007/s11023-018-9482-5
8. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2

9. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1–21. https://doi.org/10.1177/2053951716679679
10. Coeckelbergh, M. (2020). *AI Ethics*. MIT Press.
11. Sætra, H. S. (2021). A framework for evaluating and developing human-centered AI. *Humanities and Social Sciences Communications, 8*, 1–9.
12. Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction, 36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118
13. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., & others. (2019). Guidelines for human-AI interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300233
14. Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007
15. Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072
16. Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230–253. https://doi.org/10.1518/001872097778543886
17. Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8
18. Hancock, P. A. (2019). Some pitfalls in the promises of automation. *Ergonomics, 62*(4), 479–495. https://doi.org/10.1080/00140139.2018.1466859
19. Siau, K., & Wang, W. (2020). Artificial intelligence, machine learning, automation, robotics, and smart systems. *Journal of Database Management, 31*(1), 1–20.
20. UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO, Paris.