Theory and Practice of Science and Technology
2025, VOL. 6, NO. 12, 142-145
DOI: 10.47297/taposatWSP2633-456930.20250612

A Dialectical Perspective on Technological Alienation in the AI Era: Ethics, Trust, and Human-Centered Design

Yepeng Liu

School of Mechanical-Electronic and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, Beijing 102616, China

ABSTRACT

As artificial intelligence technologies deeply integrate into production and social systems, they simultaneously trigger ethical crises including labor alienation, cognitive degradation, and emotional detachment. This study builds upon Marxist alienation theory and Engels' proposition that "labor created humanity" to examine the dual nature of technological alienation in the AI era. We propose a theoretical framework of "technological alienation—trust crisis—synergistic reconstruction" from a human factors engineering perspective. Through comparative analysis of industrial manufacturing and service sector cases, we identify transparency design, dynamic authority allocation, and cross-cultural adaptation as critical mechanisms for balancing instrumental and value rationality. We further propose a human-centered AI development paradigm encompassing technical, organizational, and policy dimensions. Our findings suggest that while technological alienation cannot be eliminated, it can be dynamically managed through dialectical reconciliation. This research contributes theoretical foundations for ethical governance in human-AI collaboration and practical guidelines for human-centered technology design.

KEYWORDS

Technological alienation; Human factors engineering; AI ethics; Trust calibration; Human-centered design

1 Introduction

The Fourth Industrial Revolution, characterized by artificial intelligence, Internet of Things, and big data analytics, has fundamentally transformed production relations and social structures (Reiman et al., 2021). Marx's theory of labor alienation from the *Economic and Philosophic Manuscripts of 1844* manifests in new forms within this context, as algorithmic management systems convert human labor into quantifiable data streams, potentially eroding worker autonomy (Marx, 1844/1985). Empirical research demonstrates that in human-robot interaction scenarios, humans tend to attribute failures to machines while claiming credit for successes—a phenomenon revealing technology's subtle erosion of human agency (Lei & Rau, 2023).

Beyond production settings, algorithmic dependence has permeated daily life. Navigation systems, medical diagnostic AI, and recommendation algorithms increasingly shape human decision-making, potentially diminishing critical thinking capabilities (Glikson & Woolley, 2020). Overreliance on autonomous driving systems, for example, can impair drivers' situational awareness, creating what researchers term "automation dependency syndrome" (Walker et al., 2018). Furthermore, the proliferation of human-machine interaction is gradually replacing traditional interpersonal collaboration, raising concerns about emotional detachment in increasingly digitized societies.

The traditional paradigm of Human Factors and Ergonomics (HF/E), centered on efficiency optimization, requires fundamental reconceptualization to address ethical challenges in the AI era (Reiman et al., 2021). This study argues for a paradigm shift from "instrumental rationality" to "value rationality," positioning human factors engineering as a crucial mediator between technological advancement and human values. This reconceptualization operates across three dimensions: technical design prioritizing human-centered principles, labor processes redesigned to preserve human autonomy, and human factors engineering addressing technology's impact on social relationships.

2 Theoretical Framework: The Duality of Technological Alienation

Technological alienation embodies inherent contradictions between progress and human costs. Marx observed in *Capital* that industrial machinery simultaneously enhances productivity while reducing workers to "appendages of machines" (Marx, 1867/1992). This dialectical relationship manifests distinctly in the AI era:

Positive Alienation: Efficiency RevolutionAI technologies significantly enhance productivity across domains. For example, manufacturing facilities implementing robotic automation report 90% reduction in labor costs alongside 300% productivity increases. Human-robot collaborative systems improve task accuracy by 15-20% in industrial settings (Gualtieri et al., 2021). This efficiency revolution aligns with Engels' observation that "labor is the source of all wealth and culture" (Engels, 1876/1955).

Negative Alienation: Erosion of Human ValueEfficiency gains often come at the cost of fragmented labor processes and diminished human agency. Human-robot interaction experiments reveal that in competitive scenarios, humans frequently surrender decision-making authority to AI systems, creating "responsibility hollowing" (Lei & Rau, 2023). This alienation manifests as cognitive degradation, emotional detachment, and loss of meaning in labor.

Confronted with technological alienation's dual nature, human factors engineering must transcend traditional efficiency paradigms to fulfill dual missions of "technological critique" and "human reclamation." This dialectical role includes transparency design to counter "black box" effects and task autonomy allocation to reclaim labor subjectivity. Human factors engineering's ultimate objective is not eliminating technological alienation but achieving dynamic balance between human and machine capabilities.

3 Contradictions in Human Factors Practice

Explainable AI (XAI) approaches mitigate alienation by enhancing system transparency. Riedl's (2019) work demonstrates how AI systems generating human-like explanations help users understand algorithmic reasoning. Medical AI systems using Grad-CAM visualization allow physicians to verify whether algorithmic attention aligns with clinical expertise (Selvaraju et al., 2017).

However, pursuing maximum explainability risks "cognitive overload." Complex technical explanations may exceed users' information processing capacity. For example, displaying sensor fusion mathematics in real-time to non-technical drivers could cause decision paralysis. Effective design requires balancing explanation depth with cognitive load through tiered explanation strategies.

Anthropomorphic design in collaborative robots enhances emotional trust through simulated empathy. Research confirms that physical embodiment and social characteristics significantly increase emotional trust in AI systems (Glikson & Woolley, 2020). Nevertheless, anthropomorphic design presents "emotional exploitation" risks. Customer service chatbots simulating empathy may foster unhealthy emotional dependence while lacking genuine emotional capacity (Yan, 2019).

Human factors optimizations can inadvertently conceal algorithmic bias. For example, recruitment AI systems with interface designs emphasizing "fairness and objectivity" may mask gender discrimination embedded in training data (Barocas & Selbst, 2016). Similarly, efficiency-oriented designs in manufacturing collaboration systems may reinforce cognitive biases against older workers.

The tension between mitigating and exacerbating alienation reflects fundamental conflicts between instrumental and value rationality. Addressing these dynamic contradictions requires adaptive design frameworks including real-time feedback mechanisms, ethical auditing systems, and user empowerment design (EU AI Act, 2023).

4 Case Studies: Industrial and Service Applications

A German automotive manufacturing facility implemented a human-centered approach to human-robot collaboration. Engineers designed a system where workers maintained control over complex decision-making, while robots handled repetitive tasks. This task allocation preserved worker autonomy and prevented skill atrophy (Lei & Rau, 2023).

In stark contrast, Foxconn's fully automated "lights-out factory" eliminated 90% of assembly line workers, increasing productivity by 300%. However, remaining workers were relegated to "button operator" roles with minimal decision authority. This implementation exemplifies Marx's prediction of labor alienation: workers became disconnected from production processes, experiencing skill degradation and heightened "replaceability anxiety."

A major hospital implemented an AI-assisted diagnostic system that reduced case analysis time from 30 minutes to 5 minutes. However, patient satisfaction declined as they perceived physicians as increasingly dependent on algorithmic recommendations without personal engagement. Research indicates that 68% of patients distrust medical AI due to perceived lack of human compassion (Glikson & Woolley, 2020).

A more balanced approach emerged in a healthcare network's "physician-review AI recommendation" hybrid system. The AI provided preliminary diagnostic suggestions with Grad-CAM visualizations highlighting key image features (Selvaraju et al., 2017), while physicians retained final decision authority. This design incorporated Riedl's (2019) explainable AI principles while implementing dynamic authority allocation. Results showed a 12% improvement in diagnostic accuracy with patient trust recovering to 95% of traditional consultation levels.

5 Human-Centered AI Development Framework

5.1 Technical Dimension: Transparent and Adaptive Systems

The EU AI Act (2023) mandates "explainability statements" for high-risk AI systems, establishing a regulatory foundation for transparency design. Riedl's (2019) "automated rationale generation" transforms AI decision processes into human-understandable narratives. For example, medical AI might explain: "I recommend biopsy based on the irregular margins observed in your lung nodule on CT imaging."

Lee and See's (2004) trust calibration theory suggests human reliance on AI should correspond to task risk levels. Autonomous driving systems exemplify this principle: high-risk urban environments employ "human-dominated, AI-assisted" modes, while lower-risk highway scenarios allow greater AI autonomy. An automotive manufacturer's L3 autonomous system demonstrated this approach, actively requesting human takeover with 3-second warnings when detecting complex traffic conditions, reducing accidents by 42% (Walker et al., 2018).

5.2 Organizational Dimension: Skill Development and Cultural Adaptation

Germany's "Industry 4.0" worker retraining initiative integrates human factors principles into industrial transformation. At Volkswagen factories, workers complete training in "human-AI collaborative interface design" and "algorithmic bias recognition," enabling meaningful participation in technical decisions (Reiman et al., 2021).

Research indicates significant cultural differences in human-AI interaction preferences. Cross-cultural studies found collectivist cultures favor collaborative AI designs, while individualist cultures prefer autonomous technologies. A multinational corporation adapted its customer service chatbots accordingly, increasing global user satisfaction by 28% (Tolzin & Janson, 2023).

5.3 Policy Dimension: Multi-Stakeholder Governance

Following Reiman et al.'s (2021) "macro human factors" framework, multi-stakeholder ethics review committees can prevent algorithmic harm. France now requires patient representatives in medical AI validation processes, ensuring systems adhere to "do no harm" principles. This collaborative approach successfully identified hidden discrimination against persons with disabilities in a recruitment AI system before deployment (Barocas & Selbst, 2016).

Canada's Pan-Canadian AI Strategy incorporates "technological alienation compensation" into labor regulations, requiring companies to provide retraining or transition support for workers displaced by automation. This policy reflects Marx's insight that social institutions must mitigate labor alienation, rebuilding worker agency through systemic protections.

6 Conclusion

This study examined technological alienation in the AI era through a human factors engineering lens, yielding two primary conclusions:

The Inevitability and Manageability of Technological AlienationTechnological alienation emerges as an inherent feature of technical progress, rooted in the perpetual tension between instrumental rationality and human values. Comparative analysis reveals that while technological alienation cannot be eliminated, human factors engineering's "adaptive design" approach can achieve dynamic balance. German automotive manufacturing's collaborative human-robot systems demonstrated how optimized task allocation preserves worker skills while enhancing productivity, validating the dialectical unity of "alienation-repair."

Human Factors Engineering as an "Alienation Repair Tool"Human factors engineering mitigates technological alienation through transparency design, dynamic authority allocation, and contextual adaptation. The medical hybrid decision system case demonstrated how explainable AI combined with responsibility-sharing mechanisms effectively reconciles efficiency and trust. This role realization follows Liu's (2009) technology value theory by embedding human values within instrumental rationality, ensuring technological development serves human wellbeing.

Future research should focus on longitudinal trust dynamics, cross-cultural ethical frameworks, explainability boundaries, and quantitative alienation assessment. Technological alienation in the Industry 4.0 era represents both the price of human civilization's advancement and an opportunity for social innovation. This study has constructed a dialectical framework for understanding technological alienation and proposed practical pathways for human-centered technology development. Future research must continue exploring the theoretical and practical dimensions of human-AI collaboration to ensure AI technologies become instruments of human liberation rather than new mechanisms of alienation.

About the Author

Yepeng Liu, Master's degree candidate, Research Interests: Human factors, human-AI interaction.

References

[1] Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact[J]. *California Law Review, 104*, 671-732.

[2] Engels, F. (1955). *Dialectics of nature* (Clemens Dutt, Trans.)[M]. Foreign Languages Publishing House. (Original work published 1876).

[3] EU AI Act. (2023). *Regulation on a European approach for Artificial Intelligence[Z]*. European Parliament and Council.

[4] Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research[J]. *Academy of Management Annals, 14* (2), 627-660.

[5] Gualtieri, L., Rauch, E., & Vidoni, R. (2021). Emerging research fields in safety and ergonomics in industrial collaborative robotics: A systematic literature review[J]. *Robotics and Computer-Integrated Manufacturing, 67*, 101998.

[6] Lei, X., & Rau, P.-L. P. (2023). Effects of task structure and attribution target on responsibility attributions in human – robot interaction[J]. *International Journal of Human-Computer Studies, 176*, 103038.

[7] Marx, K. (1985). *Economic and philosophic manuscripts of 1844* (M. Milligan, Trans.)[M]. Prometheus Books. (Original work published 1844).

[8] Marx, K. (1992). *Capital: A critique of political economy* (Vol. 1). Penguin Classics. (Original work published 1867)

[9] Reiman, J., Kaivo-oja, E., Parviainen, E.-P., & Takala, E.-P. (2021). Human factors and ergonomics in manufacturing in the industry 4.0 context – A scoping review[J]. *Technology in Society, 65*, 101572.

[10] Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning[J]. *Human Behavior and Emerging Technologies, 1*(1), 33-36.

[11] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]. *Proceedings of the IEEE International Conference on Computer Vision*, 618-626.

[12] Tolzin, A., & Janson, A. (2023). Mechanisms of common ground in human-agent interaction: A systematic review of conversational agent research[C].*Proceedings of the 56th Hawaii International Conference on System Sciences*, 342-351.

[13] Walker, F., Verwey, W., & Martens, M. (2018). Gaze behaviour as a measure of trust in automated vehicles[C]. *Proceedings of the 6th Humanist Conference*, 1-6.

[14] Yan, H. (2019). Decoding AI ethics through trust[J]. *Journal of Artificial Intelligence, 4*, 95-101.