Behavioral Responses to AI-Generated Performance Targets in Flexible Work Environments

Henrik J. Gregory
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
gregory985@colostate.edu

Brooks Kelley
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
brooksk@unh.edu

Aapo Lane
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
aapolane78@buffalo.edu

Abstract

The proliferation of artificial intelligence in organizational management has introduced a new paradigm for setting performance targets, particularly within flexible work environments where traditional supervisory mechanisms are attenuated. This paper investigates the behavioral responses of workers to AI-generated performance targets, conceptualizing these targets as emergent properties of complex socio-technical systems rather than as neutral optimization tools. We argue that the deployment of algorithmic target-setting architectures fundamentally alters the psychological contract between workers and organizations, shifting from negotiated, human-mediated goal structures to dynamically computed, data-driven benchmarks. Drawing on a synthesis of systems theory, behavioral economics, and organizational psychology, we develop a framework that examines the structural trade-offs inherent in such systems, including the tension between algorithmic efficiency and worker autonomy, the robustness of target generation models under distributional shift, and the fairness implications of opaque performance metrics. We analyze how flexible work environments amplify these dynamics by decoupling temporal and spatial supervision, thereby increasing reliance on algorithmic proxies for productivity. The paper further explores governance and policy implications, including the need for transparency standards, audit mechanisms for algorithmic bias, and participatory design approaches that incorporate worker voice into target-setting processes. Case illustrations from platform-mediated gig work and remote knowledge work are used to ground the discussion. We conclude by outlining a research agenda focused on the long-term sustainability of AI-driven performance management systems, emphasizing the importance of interdisciplinary approaches to ensure that such systems remain both efficient and equitable.

Keywords

artificial intelligence, performance targets, flexible work, algorithmic management, socio-technical systems, behavioral responses, governance.

1. Introduction

The integration of artificial intelligence into organizational management has accelerated rapidly over the past decade, driven by advances in machine learning, ubiquitous sensor data, and the expansion of digital work platforms [1]. Among the most consequential applications of this technology is the automated generation of performance targets, which are now used to guide worker behavior across a wide range of industries, from logistics and retail to software development and professional services [2]. These AI-generated targets promise to optimize productivity by leveraging large-scale data to set goals that are theoretically calibrated to individual capabilities, market conditions, and organizational objectives. However, the behavioral consequences of replacing human-managed goal-setting with algorithmic systems remain poorly understood, particularly in flexible work environments where workers already operate with significant autonomy and reduced direct oversight [3].

Flexible work environments, characterized by temporal and spatial discretion, have become increasingly prevalent in the wake of the COVID-19 pandemic and the ongoing digitization of labor markets [4]. In such settings, traditional mechanisms of performance management, such as periodic face-to-face reviews and supervisor-mediated goal negotiation, are often replaced or supplemented by algorithmic systems that continuously compute and update targets based on real-time data streams [5]. This shift introduces a fundamental change in the nature of work: performance targets are no longer static, negotiated agreements but dynamic, computationally derived benchmarks that may adjust in response to worker behavior, system-wide demand, or algorithmic inference. The implications of this transformation are profound, affecting worker motivation, stress, trust, and ultimately the sustainability of the employment relationship itself [6].

This paper adopts a systems-level perspective to examine the behavioral responses elicited by AI-generated performance targets in flexible work environments. We argue that these responses cannot be understood solely through the lens of individual psychology or economic incentives; instead, they must be analyzed as emergent phenomena arising from the interaction between algorithmic architectures, organizational policies, and worker agency. Our analysis focuses on structural trade-offs, such as the balance between optimization and autonomy, the robustness of target-setting models to behavioral gaming, and the fairness of outcomes across heterogeneous worker populations. We also consider governance and policy implications, drawing on insights from socio-technical systems theory and critical algorithm studies to propose design principles for more equitable and resilient performance management systems.

2. Theoretical Foundations: Algorithmic Management and the Socio-Technical System

The concept of algorithmic management provides a foundational lens for understanding AI-generated performance targets. Algorithmic management refers to the use of computational systems to coordinate, evaluate, and direct labor, often replacing or augmenting human managers in these functions [7]. In flexible work environments, algorithmic management is particularly pervasive, as digital platforms collect granular data on worker activities, output, and efficiency, which are then fed into models that generate targets, allocate tasks, and determine compensation [8]. This creates a closed-loop system in which worker behavior is both measured and shaped by the same algorithmic infrastructure.

From a socio-technical systems perspective, the introduction of AI-generated targets is not a neutral technical intervention but a reconfiguration of the organizational environment that affects all components of the system, including workers, managers, technology, and institutional rules [9]. Socio-technical systems theory emphasizes that technical and social subsystems are interdependent and must be jointly optimized to achieve system performance and worker well-being. When AI-generated targets are imposed without adequate consideration of the social subsystem, unintended consequences such as gaming, burnout, or resistance are likely to emerge [10]. For example, workers may learn to manipulate the metrics used to generate targets, a phenomenon known as Goodhart's law in economics, which states that when a measure becomes a target, it ceases to be a good measure [11].

Behavioral economics and organizational psychology offer additional theoretical resources for understanding how workers respond to algorithmically set goals. Goal-setting theory, a well-established framework, posits that specific, challenging goals enhance performance, provided that individuals accept the goals and receive feedback on progress [12]. However, this theory was developed in contexts where goals are set through social processes involving negotiation, commitment, and mutual understanding. AI-generated targets may lack the legitimacy and transparency that underpin goal acceptance, particularly if workers perceive the targets as arbitrary, unfair, or unattainable [13]. Moreover, the dynamic nature of algorithmic targets, which can change without explanation or warning, may undermine the stability needed for effective goal pursuit.

3. Structural Trade-Offs in AI-Driven Target Generation

The deployment of AI-generated performance targets involves several fundamental structural trade-offs that affect both organizational efficiency and worker experience. One key trade-off is between algorithmic precision and worker autonomy. Proponents of algorithmic management argue that data-driven targets can be more accurate than human-set goals, because they incorporate a wider range of information and are not subject to cognitive biases such as anchoring or recency effects [14]. However, the same precision can be experienced as oppressive by workers, who may feel that their performance is being micromanaged by an opaque system that does not account for contextual factors or personal circumstances. In flexible work environments, where autonomy is a primary source of job satisfaction, the erosion of self-determination through algorithmic target-setting can lead to decreased motivation and increased turnover [15].

A second trade-off concerns the robustness of target-generation models under distributional shift. Machine learning models are typically trained on historical data that reflect past patterns of work behavior, market conditions, and organizational priorities. When these patterns change, as they frequently do in dynamic labor markets, the models may produce targets that are misaligned with current realities [16]. For instance, a model trained on pre-pandemic data may generate unrealistic targets for remote workers facing new constraints. Moreover, the feedback loop between worker behavior and target adjustment can amplify initial errors, leading to a phenomenon known as algorithmic volatility, where targets fluctuate excessively in response to short-term variations in performance [17]. This volatility can be particularly damaging in flexible work settings, where workers already contend with income instability and irregular schedules.

A third trade-off involves fairness and equity. AI-generated targets may perpetuate or exacerbate existing disparities if the underlying data reflect biased historical practices or if the model's features are correlated with protected characteristics such as race, gender, or age [18]. For example, a target-setting algorithm that uses past productivity as a predictor may penalize workers who have taken parental leave or who work part-time, even if their current capacity is equal to that of full-time workers. In flexible work environments, where workers often have diverse schedules and responsibilities, the risk of algorithmic discrimination is heightened. Furthermore, the opacity of many AI systems makes it difficult for workers to challenge or even understand how their targets are set, undermining procedural justice and trust [19].

4. Behavioral Responses: Adaptation, Gaming, and Resistance

Workers in flexible environments exhibit a range of behavioral responses to AI-generated performance targets, which can be categorized along a spectrum from adaptation to gaming to outright resistance. Adaptation refers to behaviors that align with the algorithmic system's intended goals, such as adjusting work patterns to meet targets or acquiring new skills to improve performance. While adaptation can be beneficial for both workers and organizations, it may come at a cost if workers sacrifice long-term well-being for short-term target attainment [20]. For instance, gig workers on delivery platforms may rush to complete orders to meet algorithmic targets, increasing the risk of accidents or burnout.

Gaming behaviors involve strategic actions designed to manipulate the metrics used by the algorithm without necessarily improving actual productivity. Common examples include working only during periods when the algorithm assigns higher rewards, artificially inflating output metrics, or coordinating with other workers to exploit system rules [21]. Gaming is a rational response to systems that are perceived as unfair or opaque, but it undermines the validity of the performance data and can lead to suboptimal organizational outcomes. In flexible work environments, where supervision is minimal, the opportunities for gaming are amplified, and the costs of detection are low.

Resistance encompasses more overt forms of opposition, such as collective bargaining for algorithmic transparency, refusal to comply with target demands, or exit from the platform or organization altogether [22]. Resistance is more likely when workers perceive that the algorithmic system violates their sense of dignity or fairness, or when they have access to alternative employment opportunities. The rise of worker-led movements advocating for algorithmic accountability, such as the calls for transparency in Uber's driver rating system, illustrates the growing recognition of the power asymmetries inherent in AI-managed work [23].

5. Governance and Policy Implications

The behavioral responses described above have significant implications for the governance of AI-generated performance targets in flexible work environments. A purely technocratic approach that seeks to optimize target-setting algorithms without addressing the social and organizational context is unlikely to succeed in the long term. Instead, governance frameworks must be developed that incorporate principles of transparency, accountability, and participation.

Transparency is a necessary condition for trust in algorithmic systems. Workers should have access to information about how their targets are computed, what data are used, and how the models are updated [24]. However, transparency alone is insufficient if the underlying algorithms are too complex for workers to understand. Therefore, organizations should also provide meaningful explanations of target-setting decisions, perhaps through interpretable model summaries or human-readable justifications. Regulatory bodies may need to establish standards for algorithmic transparency in employment contexts, similar to the European Union's General Data Protection Regulation provisions on automated decision-making.

Accountability mechanisms are equally important. When algorithmic targets lead to adverse outcomes, such as unfair treatment or excessive stress, there must be avenues for workers to challenge and seek redress. This could involve internal grievance procedures, external audits by independent third parties, or regulatory oversight [25]. The design of accountability systems should account for the power imbalances between workers and platform owners, ensuring that workers have the resources and support needed to exercise their rights.

Participatory design approaches offer a promising pathway for aligning algorithmic systems with worker values and needs. By involving workers in the development and refinement of target-setting models, organizations can benefit from local knowledge about work practices, contextual constraints, and fairness considerations [26]. Participatory processes also enhance the legitimacy of algorithmic decisions, reducing the likelihood of resistance and gaming. In flexible work environments, where workers are geographically dispersed and often lack formal representation, digital participatory mechanisms such as online surveys, feedback forums, and worker councils can facilitate meaningful input.

6. Case Illustrations: Gig Work and Remote Knowledge Work

Two illustrative cases highlight the dynamics discussed in this paper. The first is platform-mediated gig work, such as ride-hailing or food delivery, where AI-generated targets are used to assign tasks, set earnings goals, and determine bonuses. Research has shown that gig workers often respond to these targets by altering their work schedules, accepting or rejecting tasks strategically, and sometimes coordinating with peers to game the system [5]. The dynamic and opaque nature of targets in these platforms contributes to worker anxiety and a sense of precarity, as workers cannot predict how their behavior will affect future opportunities.

The second case is remote knowledge work, such as software development or consulting, where AI-generated targets may be used to set project milestones, billable hour goals, or code quality metrics. In these settings, workers have greater autonomy and cognitive complexity, but algorithmic target-setting can still lead to negative outcomes if it fails to account for the collaborative and creative nature of knowledge work. For example, a target based on lines of code may incentivize quantity over quality, while a target based on task completion time may discourage thoroughness and learning. The behavioral responses in knowledge work often involve subtle forms of resistance, such as ignoring algorithmic recommendations or developing workarounds that preserve professional autonomy.

7. Future Directions and Research Agenda

The study of behavioral responses to AI-generated performance targets is still in its early stages, and many questions remain unanswered. Future research should investigate the long-term effects of algorithmic target-setting on worker health, career trajectories, and organizational culture. Longitudinal studies that track workers over multiple years could reveal whether the initial adaptation to algorithmic targets gives way to burnout or disengagement, or whether workers develop effective coping strategies.

Another important direction is the development of fairness-aware algorithms that can detect and mitigate bias in target-setting. Researchers in algorithmic fairness have made progress in areas such as credit scoring and criminal justice, but the application of these techniques to performance management presents unique challenges, including the need to define fairness in contexts where productivity is itself a contested construct [27]. Interdisciplinary collaborations between computer scientists, organizational psychologists, and labor economists will be essential for designing systems that are both efficient and equitable.

Finally, the policy implications of AI-driven performance management warrant deeper investigation. As governments around the world consider regulations for algorithmic decision-making in employment, empirical evidence on the behavioral and economic impacts of different governance approaches will be critical. Comparative studies of jurisdictions with varying levels of algorithmic transparency and worker protections could provide valuable insights into best practices.

8. Conclusion

AI-generated performance targets represent a significant shift in the management of flexible work, with far-reaching implications for worker behavior, organizational efficiency, and social equity. This paper has argued that these targets must be understood as embedded within complex socio-technical systems, where technical design choices interact with human psychology and organizational governance to produce emergent behavioral outcomes. The structural trade-offs between precision and autonomy, robustness and volatility, and efficiency and fairness are inherent to algorithmic target-setting and cannot be resolved through purely technical means. Instead, organizations and policymakers must adopt a holistic approach that prioritizes transparency, accountability, and participatory design. By doing so, they can harness the benefits of AI-driven performance management while mitigating its risks, ensuring that the future of work is not only more productive but also more just and sustainable.

References

1. Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. Academy of Management Annals, 14(1), 366-410.

2. Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. Work, Employment and Society, 33(1), 56-75.

3. Jarrahi, M. H., & Sutherland, W. (2019). Algorithmic management and algorithmic competencies: Understanding and appropriating algorithms in gig work. In Information in Contemporary Society (pp. 578-589). Springer.

4. Choudhury, P., Foroughi, C., & Larson, B. (2021). Work-from-anywhere: The productivity effects of geographic flexibility. Strategic Management Journal, 42(4), 655-683.

5. Min, X., Chi, W., Hu, X., & Ye, Q. (2024). Set a goal for yourself? A model and field experiment with gig workers. Production and Operations Management, 33(1), 205-224.

6. Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 1603-1612).

7. Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. Human Resource Management Journal, 30(1), 114-132.

8. Rosenblat, A., & Stark, L. (2016). Algorithmic labor and information asymmetries: A case study of Uber's drivers. International Journal of Communication, 10, 3758-3784.

9. Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting. Human Relations, 4(1), 3-38.

10. Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

11. Strathern, M. (1997). 'Improving ratings': Audit in the British university system. European Review, 5(3), 305-321.

12. Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American Psychologist, 57(9), 705-717.

13. Bhave, D. P., Teo, L. H., & Dalal, R. S. (2020). Privacy at work: A review and a research agenda for a contested terrain. Journal of Management, 46(1), 127-164.

14. Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction machines: The simple economics of artificial intelligence. Harvard Business Review Press.

15. Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. Journal of Organizational Behavior, 26(4), 331-362.

16. Quinonero-Candela, J., Sugiyama, M., Schwaighofer, A., & Lawrence, N. D. (2009). Dataset shift in machine learning. MIT Press.

17. Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-14).

18. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671-732.

19. Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. Journal of Applied Psychology, 86(3), 425-445.

20. Caza, A., & Carroll, S. J. (2012). Critical theory and organizational behavior. In The Oxford handbook of organizational psychology (pp. 123-145). Oxford University Press.

21. Bernstein, E. S. (2012). The transparency trap: How visibility can undermine learning and performance. Harvard Business Review, 90(10), 99-105.

22. Anwar, M. A., & Graham, M. (2021). Between a rock and a hard place: The role of data and algorithms in the gig economy. Big Data & Society, 8(1), 1-14.

23. Ticona, J., & Mateescu, A. (2018). Trusted strangers: Carework platforms' cultural entrepreneurship in the on-demand economy. New Media & Society, 20(11), 4384-4404.

24. Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. University of Pennsylvania Law Review, 165(3), 633-705.

25. Selbst, A. D., & Powles, J. (2017). Meaningful information and the right to explanation. International Data Privacy Law, 7(4), 233-242.

26. Muller, M., & Druin, A. (2012). Participatory design: The third space in HCI. In The human-computer interaction handbook (pp. 1125-1154). CRC Press.

27. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (pp. 214-226).