DOCTORAL CONSORTIUM

# STRATEGIC ADOPTION OF AI-ENABLED DECISION-MAKING SYSTEMS: DESIGNING FOR HUMAN AGENCY IN DATA-DRIVEN ORGANIZATIONS

BARRAK ALBABTAIN

University College Dublin, UCD Michael Smurfit Graduate Business School, Dublin, Ireland
barrak.albabtain@ucdconnect.ie

**DOI**
https://doi.org/10.18690/um.fov.4.2026.72

**ISBN**
978-961-299-152-4

Artificial intelligence is increasingly embedded in organizational decision-making, reshaping how managers exercise discretion and responsibility. This doctoral research examines how AI-enabled decision systems affect human agency in data-driven organizations. It focuses on how technological design features, including transparency and override flexibility, interact with governance structures such as accountability and incentive systems. Using a sequential multi-phase design combining experiments and qualitative fieldwork, the study investigates both perceived and enacted managerial agency. The intended contribution is an Information Systems framework explaining when AI supports human augmentation and when it produces functional substitution. The research aims to inform the design of responsible sociotechnical systems that preserve meaningful human involvement while retaining the efficiency benefits of AI-enabled decision support.

**Keywords:**
artificial intelligence;
human-AI interaction;
decision support systems;
algorithmic governance;
managerial discretion;
information systems design

University of Maribor Press

## 1 Introduction

Artificial intelligence is increasingly embedded in organizational decision systems across hiring, performance evaluation, financial risk assessment, forecasting, and resource allocation (Faraj et al., 2018; Kellogg et al., 2020). Organizations adopt AI-enabled systems to improve efficiency, scalability, and predictive accuracy. However, these systems do more than support decisions. They shape how decisions are structured, justified, and governed (Shrestha et al., 2019; Raisch & Krakowski, 2021).

Information Systems research has examined algorithmic transparency, explainability, trust calibration, bias mitigation, and human-AI collaboration (Glikson & Woolley, 2020; Burton et al., 2020; Jarrahi, 2018). While these perspectives are valuable, much existing work treats AI mainly as an input into decision-making rather than as a structural component of decision systems. When AI systems generate recommendations, rankings, or classifications, they introduce new forms of epistemic authority that may alter how managerial discretion is exercised (Berente et al., 2021).

Recent AI governance developments intensify this issue. The European Union Artificial Intelligence Act establishes risk-based requirements for transparency, accountability, and human oversight in high-impact AI systems (European Union, 2024; Cancela-Outeda, 2024). Organizational responsible AI frameworks similarly emphasize documentation, auditability, and structured oversight (Papagiannidis et al., 2025). These governance arrangements do not simply constrain AI use. They shape how decision systems are designed and how authority is distributed between humans and algorithms.

This doctoral research asks how AI-enabled decision systems reshape human agency in organizational contexts. Human agency is understood as the capacity to exercise meaningful discretion, interpret information, and assume responsibility for outcomes. The project examines when AI functions as augmentation, preserving interpretive authority, and when it becomes functional substitution, where human actors formally approve algorithmic outputs while exercising limited substantive discretion.

## 2 Problem Definition

Much of the current literature focuses on AI adoption, trust, fairness, or performance outcomes. Less attention has been given to how AI integration reshapes managerial discretion itself. This gap matters because managers may remain formally accountable for decisions even when algorithmic outputs strongly shape what choices appear legitimate, rational, or defensible.

Human-in-the-loop arrangements are often treated as sufficient to preserve agency. However, the mere presence of a human actor in a workflow does not guarantee meaningful discretion. Override mechanisms may exist but be procedurally burdensome. Performance metrics may reward adherence to algorithmic outputs. Accountability structures may diffuse responsibility and discourage deviation. In these situations, substitution can emerge indirectly through system design and governance architecture rather than explicit automation.

This research therefore conceptualizes AI integration along an augmentation-substitution continuum (Raisch & Krakowski, 2021). Augmentation refers to configurations in which AI expands informational capacity while preserving meaningful human interpretation and override authority. Substitution refers to configurations in which AI effectively determines outcomes, with humans serving primarily as validators of system outputs.

The central research question is: How do AI-enabled decision systems reshape human agency in organizational contexts, and under what conditions do they function as augmentation rather than substitution?

The study is guided by three sub-questions. First, how do design features such as algorithmic transparency and override flexibility influence perceived and enacted managerial agency? Second, how do accountability structures and incentive systems shape the legitimacy and frequency of deviation from AI recommendations? Third, how do managers interpret and enact discretion in AI-supported strategic decisions?

## 3 Methodology

This doctoral research adopts a sequential multi-phase design. The phased approach allows the study to move from controlled mechanism testing to contextualized organizational understanding and finally to integrative theory development. The design consists of three phases: an experiment on perceived agency, qualitative fieldwork on enacted agency, and integrative model development.

Phase 1 uses experimental methods to identify causal mechanisms linking AI system design features to perceived managerial agency. Participants will assume managerial roles in simulated decision environments and evaluate cases in which AI systems provide recommendations. The experiment will manipulate algorithmic transparency, strength of recommendation, override flexibility, and accountability framing. Dependent variables will include perceived control, responsibility attribution, confidence in decision quality, and likelihood of override.

The purpose of Phase 1 is to strengthen internal validity by isolating how specific design features influence agency-related outcomes. For example, transparency may increase perceived control by making algorithmic outputs more interpretable, while high-friction override procedures may reduce willingness to deviate even when participants disagree with the recommendation.

Phase 2 uses qualitative case studies to examine enacted agency in organizations implementing AI-enabled decision systems. Data collection will include semi-structured interviews with managers, analysts, and governance officers. Where possible, it will also include policy documents, workflow diagrams, audit procedures, and decision logs. The qualitative phase will examine how managers interpret algorithmic outputs, when they override or follow recommendations, and how organizational norms shape the legitimacy of deviation.

Phase 3 integrates findings from the experimental and qualitative phases into a conceptual framework. This framework will link technological design features and institutional governance structures to perceived and enacted agency outcomes. The sequential logic is important: the experiment provides mechanism-level clarity, while fieldwork examines how those mechanisms operate within real organizational settings.

The design explicitly addresses validity. Internal validity is supported through controlled manipulation in Phase 1. Construct validity is strengthened by distinguishing perceived agency from enacted agency. External validity is supported through field-based evidence from organizational contexts. Triangulation across experimental and qualitative data also reduces reliance on a single source of evidence.

## 4 Preliminary and Expected Results

Although data collection is not yet complete, the research anticipates several patterns. First, algorithmic transparency is expected to increase perceived agency by enabling managers to interpret recommendations rather than merely receive them (Bansal et al., 2021). However, transparency alone may not ensure enacted discretion if organizational norms discourage deviation from algorithmic outputs.

Second, override flexibility is expected to shape enacted agency. When deviation from AI recommendations is easy, legitimate, and low-friction, managers may be more willing to integrate contextual judgment. When override requires extensive documentation, approval, or exposes managers to scrutiny, they may default to algorithmic recommendations despite reservations. This suggests that substitution may arise from procedural design rather than direct automation.

Third, accountability structures are expected to moderate the relationship between AI design and agency. If managers remain clearly responsible for outcomes, they may engage more actively with AI outputs. If responsibility is diffused or attributed to algorithmic objectivity, managers may defer to system recommendations. This implies that responsible AI governance requires more than formal oversight; it requires governance structures that preserve legitimate discretion.

The expected theoretical contribution is a refined understanding of the augmentation-substitution continuum. Rather than treating augmentation or substitution as inherent properties of AI, the research argues that they are sociotechnically constructed through interactions between design, governance, and managerial practice. The distinction between perceived and enacted agency is also expected to be important, since managers may feel included in decision loops while still having limited practical discretion.

Practically, the research is expected to inform the design of AI-enabled decision systems that preserve meaningful human involvement. Potential implications include explanation interfaces that support interpretation, low-friction override pathways, explicit accountability mapping, and performance metrics that recognize contextual judgment rather than rigid adherence to algorithmic outputs.

## 5 Future Development

The next stage of this doctoral research is to finalize the experimental design and conduct pilot testing. Pilot work will refine the transparency manipulation, calibrate override conditions, and strengthen the measurement of perceived control and responsibility attribution. The main experiment will then form the basis of the first dissertation paper.

Following the experimental phase, the project will move toward organizational fieldwork. The field phase will require access to organizations using AI-enabled decision systems in areas such as hiring, performance evaluation, financial forecasting, or resource allocation. Theoretical sampling will be used to capture variation in governance structures and override protocols.

In the medium term, findings from the experimental and field phases will be integrated into a framework linking AI design, governance, perceived agency, and enacted discretion. This integrative work is intended to form a third dissertation paper and to support journal submissions in Information Systems and digital innovation outlets. Over the longer term, the research may be extended to compare regulatory or industry contexts and to develop prescriptive design guidance for responsible AI implementation.

Overall, the research trajectory moves from mechanism identification to contextual explanation and then to integrative theory building. This progression is designed to ensure both conceptual rigor and practical relevance within the doctoral timeframe.

## 6 Conclusion

This doctoral project contributes to Information Systems research by clarifying how technological design and institutional governance jointly shape human agency in AI-enabled decision systems. By combining experimental and field-based approaches, it aims to move beyond generic calls for human-in-the-loop systems toward a structured understanding of meaningful human involvement. The central argument is that AI does not automatically augment or substitute human judgment. Instead, these outcomes depend on how systems are designed, governed, and enacted in organizational practice.

**References**

Bansal, G., Wu, M., Zhou, J., Fok, R., & Kamar, E. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary human-AI performance. Proceedings of the ACM Conference on Human Factors in Computing Systems, 1-16. https://doi.org/10.1145/3411764.3445717

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. MIS Quarterly, 45(3), 1433-1450. https://doi.org/10.25300/MISQ/2021/16274

Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. Journal of Behavioral Decision Making, 33(2), 220-239. https://doi.org/10.1002/bdm.2155

Cancela-Outeda, C. (2024). The EU's AI Act: A framework for collaborative governance. Technology in Society, 78, 102565. https://doi.org/10.1016/j.techsoc.2024.102565

European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. Information and Organization, 28(1), 62-70. https://doi.org/10.1016/j.infoandorg.2018.02.005

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627-660. https://doi.org/10.5465/annals.2018.0057

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

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. https://doi.org/10.5465/annals.2018.0174

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90-103. https://doi.org/10.1016/j.obhdp.2018.12.005

Meijerink, J., Boons, M., Keegan, A., & Marler, J. (2021). Algorithmic human resource management: Synthesizing developments and cross-disciplinary insights on digital HRM. The International Journal of Human Resource Management, 32(12), 2545-2562. https://doi.org/10.1080/09585192.2021.1925326

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. Journal of Strategic Information Systems, 34(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

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

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. California Management Review, 61(4), 66-83. https://doi.org/10.1177/0008125619862257