



Leadership beyond the Individual Mind: Examining Artificial Intelligence, External Cognitive Influence, Decision-Making, and the Exercise of Power over Others

Paul Andrew Bourne¹, Dr.PH., PhD.

¹Assistant Professor, University of the Commonwealth Caribbean [UCC], Kingston, Jamaica, WI.

Abstract

Background: The increasing integration of artificial intelligence (AI), algorithmic systems, advisers, and digital decision-support technologies into organisational leadership challenges conventional assumptions that cognition, judgement, and moral responsibility reside solely within the leader. This study develops the concept of the outside mind to describe the network of external human, technological, informational, and institutional resources incorporated into leadership cognition and decision-making.

Objective: The current study seeks to develop and refine the Outside Mind Theory of Leadership by examining how external cognitive influence shapes leadership judgement, decision-making autonomy, moral agency, organisational power, and potentially oppressive leadership outcomes.

Methods: A PRISMA 2020-guided systematic qualitative review and theory-building synthesis was undertaken. Relevant peer-reviewed studies were identified through systematic database searches using predefined eligibility criteria, screened at title, abstract, and full-text stages, and critically appraised using design-appropriate tools. Extracted findings were analysed through qualitative thematic synthesis, constant comparison, and iterative theoretical integration to identify mechanisms, contextual conditions, and relationships linking cognitive augmentation, cognitive dependency, moral disengagement, employee agency, and organisational power.

Expected contribution: This study proposes that external cognition is not inherently oppressive; rather, harmful outcomes may emerge when cognitive dependency converges with weak accountability, limited transparency, moral disengagement, concentrated power, and reduced employee contestability.

Conclusion: The review provides a systematically derived qualitative foundation for theorising AI-mediated leadership and for informing future empirical testing, leadership development, and

responsible AI governance that preserves human judgement, accountability, dignity, autonomy, and employee voice.

Keywords: Artificial intelligence; Cognitive dependency; External cognition; Leadership behaviour; Organisational power; Outside Mind Theory.

Introduction

Leadership theory has historically located judgement, intentionality, moral responsibility, and decision-making primarily within the leader, even when acknowledging the influence of followers and organisational context (Bass & Bass, 2008; Yukl, 2013; Burns, 1978). However, the contemporary information environment increasingly challenges this individual-centred conception because leaders routinely think and decide through networks of advisers, consultants, dashboards, databases, algorithms, and generative AI systems. Extended-mind scholarship provides a conceptual foundation for reconsidering the location of cognition by arguing that external resources can become functionally incorporated into cognitive processes when they reliably perform roles otherwise associated with internal memory or reasoning (Clark & Chalmers, 1998; Greif, 2015; Loock, 2025; Sprevak, 2009; Telakivi, 2020). Accordingly, leadership in AI-rich organisations may increasingly be understood not as cognition occurring exclusively within the leader, but as cognition distributed across human and technological systems, “the extended mind”.

Artificial intelligence intensifies this development because it does more than store or retrieve information; contemporary systems can synthesise evidence, classify individuals, forecast outcomes, recommend courses of action, evaluate performance, and increasingly automate managerial functions. The OECD defines algorithmic management as the use of software, potentially incorporating AI, to automate wholly or partly tasks traditionally undertaken by managers, while the ILO notes that such systems can organise, assign, monitor, supervise, and evaluate work (International Labour Organization [ILO], 2025; Milanez, 2025). These technologies can improve consistency, productivity, and informational capacity. The same infrastructure can alter relationships of authority by allowing organisational decisions to be produced or justified through systems whose reasoning may be difficult for workers, and sometimes leaders themselves, to scrutinise. The critical leadership question is, therefore, no longer merely whether AI assists decision-making, but how reliance on external cognition changes who effectively exercises judgement, where responsibility resides, and how organisational power is experienced by those subjected to decisions.

This study conceptualises these external cognitive influences as the outside mind, meaning the constellation of human, technological, informational, and institutional resources that becomes incorporated into a leader's processes of interpreting situations, evaluating people, constructing alternatives, and selecting or legitimising courses of action. The concept deliberately extends beyond AI because leadership has always involved external cognition through advisers, professional networks, organisational routines, expert systems, ideological frameworks, and institutional knowledge, yet generative AI introduces unprecedented scale, speed, apparent authority, and capacity for personalised decision support. The distinction between cognitive

augmentation and cognitive dependency is therefore central: augmentation occurs when external intelligence expands a leader's informational and analytical capacity while critical judgement remains substantively human, whereas dependency emerges when the leader routinely defers interpretation or evaluation to the external system. This distinction matters because recent scholarship on AI-supported decision-making warns that automation bias and excessive reliance can weaken meaningful human scrutiny even when nominal human oversight remains formally present.

The relationship between the outside mind and oppression becomes particularly important when external cognition is embedded within organisational systems of surveillance, evaluation, reward, discipline, or exclusion. Algorithmic management research indicates that automated systems may influence employee scheduling, performance assessment, compensation, monitoring, and access to work opportunities, while studies have raised concerns about reduced worker autonomy, weakened employee voice, intrusive surveillance, and shifts in organisational power (Collins, 2023; Doellgast et al., 2023; Milanez, 2025). In the present study, oppression is not equated with ordinary managerial authority or unpopular decisions; rather, it refers to systematic organisational practices that diminish individuals' agency, voice, dignity, autonomy, participation, fair treatment, or meaningful access to opportunities and resources. AI can potentially magnify such practices when leaders portray algorithmically mediated decisions as objective or inevitable, thereby obscuring human choices embedded in system design, implementation, interpretation, and enforcement.

A further concern involves moral agency and responsibility, because leaders may psychologically distance themselves from harmful outcomes when decisions are recommended or generated by apparently authoritative external systems. Moral disengagement theory explains how individuals can participate in harmful conduct while reducing self-sanction through mechanisms such as displacement or diffusion of responsibility, advantageous comparison, euphemistic framing, and minimisation of consequences (Bandura, 1999). AI-mediated leadership may create novel opportunities for such disengagement when leaders rationalise coercive or inequitable decisions by asserting that “the data,” “the algorithm,” or “the system” produced the outcome, even though human actors selected the technology, objectives, thresholds, and conditions of implementation. Philosophical scholarship on artificial moral responsibility similarly cautions that attributing agency to machines does not resolve the underlying problem of locating human responsibility for technology-mediated harms.

The emerging literature therefore presents a paradox: AI can enhance leadership capacity while simultaneously enabling new forms of impersonalised control. OECD evidence recognises the potential for algorithmic management to improve efficiency and consistency but also documents concern regarding worker outcomes, while recent research suggests that intrusive algorithmic surveillance can alter employee autonomy, trust, and compliance (Milanez, 2025). More provocatively, experimental evidence has suggested that workers managed by AI may tolerate materially less favourable outcomes with weaker emotional reactions than workers receiving equivalent evaluations from human managers, raising the possibility that perceived machine impartiality can sometimes reduce the social resistance that ordinarily constrains exploitative practices. These findings indicate that the ethical implications of AI leadership cannot be reduced

to algorithmic accuracy; they require analysis of power, legitimacy, agency, accountability, and the social consequences of transferring cognitive authority to technological systems.

Despite growing scholarship on algorithmic management (Chigbu, 2026; Lippert et al., 2026; Sullivan et al., 2024), responsible AI (Atianashie et al., 2025; Baciú et al., 2026; Jafari et al., 2026), extended cognition (Favela et al., 2021; Kuchař, 2025; Wilson et al., 2025), and leadership technology, an important conceptual gap remains. Existing research largely examines AI adoption, algorithmic control, employee attitudes, human-AI collaboration, or responsible leadership separately, while comparatively little scholarship explains how externalised cognition itself may transform the leader's moral and cognitive relationship to power. Recent organisational research recognises both responsible-leadership opportunities created by AI and risks associated with automated management, but the literature remains fragmented across leadership studies, organisational behaviour, philosophy of mind, AI ethics, and labour research. The present study addresses this gap by proposing the outside mind as an integrative construct connecting external cognition, cognitive dependency, moral disengagement, leadership behaviour, and the potential exercise of oppressive organisational power.

Accordingly, the study advances a central proposition: external cognition is not inherently emancipatory or oppressive; its consequences depend on how cognitive authority and moral responsibility are distributed between leaders, external systems, and those affected by leadership decisions. Human oversight, transparency, critical thinking, AI literacy, ethical leadership, organisational accountability, and employee voice may preserve the benefits of cognitive augmentation while constraining the conversion of external intelligence into oppressive control. Conversely, opaque systems, automation bias, excessive cognitive delegation, weak accountability, and concentrated organisational power may transform the outside mind from a decision-support architecture into an instrument through which harmful leadership practices become technically legitimised and morally distanced. This framework therefore extends leadership research by shifting attention from whether leaders *use* AI to how deeply external cognitive systems become incorporated into leadership judgement and how that incorporation reshapes responsibility and power.

Research Gap

Despite rapidly expanding scholarship on artificial intelligence, algorithmic management, extended cognition, responsible leadership, and technology-mediated organisational decision-making, a significant theoretical gap remains concerning how external cognitive resources reshape leadership cognition and the exercise of power. The current scholarship has generally examined AI adoption, algorithmic control, human-AI collaboration, employee responses, moral responsibility, and organisational power as relatively separate areas of inquiry, resulting in a fragmented understanding of their interconnectedness. In particular, Extended Mind Theory (Clark & Chalmers, 1998; Greif, 2015) provides an important explanation of how cognition may extend beyond the biological individual. Research on algorithmic management demonstrates how technological systems increasingly participate in organisational monitoring, evaluation, coordination, and decision-making. However, these traditions have rarely been systematically integrated with scholarship on moral disengagement, employee agency, leadership responsibility,

and oppression. The literature lacks an integrative theoretical explanation of how externalised cognition may alter not only what leaders decide but also how they understand their autonomy, responsibility, and authority over individuals affected by those decisions.

An important gap in the current literature concerns the distinction between cognitive augmentation and cognitive dependency. External cognitive systems may enhance leadership by increasing access to information, improving analytical capacity, identifying alternatives, and supporting more informed decision-making, yet excessive reliance may simultaneously weaken critical scrutiny and independent judgement. The existing research has not sufficiently synthesised the conditions under which technologically extended cognition moves from supporting human judgement to becoming an authoritative cognitive resource upon which leaders routinely depend. This distinction is theoretically important because the ethical implications of AI-mediated leadership may depend less upon whether leaders use AI than upon how deeply technological systems become incorporated into leadership cognition and how much cognitive authority leaders subsequently surrender to those systems.

A second gap concerns the relationship between external cognition and moral agency. Moral disengagement theory demonstrates that displacement and diffusion of responsibility can enable individuals to participate in harmful conduct while reducing personal self-sanction, yet insufficient attention has been directed towards whether AI-mediated decision-making creates new mechanisms through which leaders distance themselves from the consequences of organisational decisions (Bandura, 1999). When leaders describe decisions as originating from “the algorithm,” “the data,” or “the system,” technological mediation may potentially obscure the continuing role of human judgement in selecting, interpreting, accepting, and implementing those recommendations. A systematic synthesis is required to determine how existing evidence connects technological cognitive authority, decision ownership, responsibility displacement, moral disengagement, and what this study conceptualises as algorithmic moral distancing.

A third gap in the current literature involves the relationship between AI-mediated cognition and the exercise of organisational power over others. Algorithmic systems increasingly participate in employee monitoring, scheduling, performance assessment, compensation, recruitment, disciplinary processes, and allocation of workplace opportunities, potentially altering employee autonomy and voice as well as the distribution of organisational power. The available literature has not been sufficiently integrated to explain when such systems support responsible and empowering leadership and when they may contribute to systematic reductions in employee agency, dignity, fairness, participation, and meaningful contestability. Addressing this fragmentation requires a systematic qualitative synthesis capable of connecting cognitive, ethical, technological, leadership, and power-related perspectives within a unified theoretical framework.

Research Objective

The principal objective of this study is to systematically identify, critically appraise, qualitatively synthesise, and theoretically integrate existing scholarly evidence concerning how external cognitive influences, particularly artificial intelligence and algorithmic decision-support systems, shape leadership cognition, judgement, moral agency, decision-making autonomy, and the

exercise of organisational power. Using a PRISMA 2020-guided systematic qualitative review and theory-building synthesis, the study seeks to establish how existing literature conceptualises the movement between cognitive augmentation and cognitive dependency and how these processes influence leadership responsibility and relationships with employees. Rather than empirically testing predetermined causal relationships, the study uses the systematically identified literature to examine, challenge, refine, and extend the proposed relationships among external cognitive influence, leadership judgement, moral disengagement, organisational power, and employee outcomes. The overarching purpose is to develop a systematically derived Outside Mind Theory of Leadership capable of explaining both the enabling and potentially oppressive consequences of externally mediated leadership cognition.

A secondary objective is to identify the mechanisms and contextual conditions determining whether the outside mind operates predominantly as a source of leadership augmentation or cognitive dependency. Particular attention was directed towards critical thinking, ethical leadership, AI literacy, algorithmic transparency, organisational accountability, employee voice, and decision contestability as potential protective mechanisms identified within the literature. Conversely, the review examines evidence concerning automation bias, technological opacity, responsibility displacement, surveillance, concentrated organisational authority, and limited opportunities for employee challenge as conditions potentially associated with harmful leadership outcomes. These findings were theoretically integrated rather than statistically tested, allowing the review to establish propositions and boundary conditions suitable for subsequent empirical investigation.

Research Questions

Consistent with its systematic qualitative and theory-building orientation, the study was guided by the following six research questions:

1. How does the existing scholarly literature conceptualise the role of external cognitive resources, particularly artificial intelligence and algorithmic systems, in shaping leadership judgement and decision-making autonomy?
2. Under what conditions does AI-mediated external cognition function as cognitive augmentation, and under what conditions does it develop into cognitive dependency within leadership decision-making?
3. How does the existing evidence explain the relationship among external cognitive dependency, decision ownership, moral agency, responsibility displacement, and moral disengagement in leadership?
4. How does AI-mediated leadership influence the exercise of organisational power and employees' experiences of voice, autonomy, dignity, fairness, participation, surveillance, and decision contestability?
5. What organisational, technological, ethical, and leadership conditions appear to constrain or facilitate the transformation of external cognitive influence into empowering/responsible leadership or oppressive leadership?
6. How can the systematically synthesised evidence be theoretically integrated to develop and refine an Outside Mind Theory of Leadership explaining the relationships among external

cognition, cognitive augmentation or dependency, leadership judgement, moral agency, organisational power, and employee outcomes?

Theoretical Framework

The current study requires a theoretical framework capable of explaining not only how leaders use artificial intelligence and other external resources, but also how those resources may become incorporated into leadership cognition and subsequently influence the exercise of power. No single existing leadership theory adequately captures this entire process; the study integrates Extended Mind Theory, Distributed Cognition, Social Cognitive Theory, Moral Disengagement Theory, and theories of organisational power. The central theoretical proposition is that leadership cognition does not necessarily remain confined to the biological mind of the leader when external technologies, advisers, information systems, and institutional structures become routinely involved in interpreting problems and recommending decisions. The theoretical framework therefore conceptualises leadership as a potentially distributed cognitive process whose ethical consequences depend upon the relationship among external cognitive influence, retained human judgement, moral agency, accountability, and organisational power.

Extended Mind Theory provides the principal cognitive foundation for the proposed concept of the outside mind (Clark & Chalmers, 1998; Greif, 2015; Looock, 2025; Sprevak, 2009; Telakivi, 2020). Clark and Chalmers (1998) challenged conventional assumptions that cognition necessarily ends at the boundaries of the brain by arguing that external resources may become functionally integrated into cognitive processes when individuals reliably use them to accomplish cognitive tasks. Applied to leadership, an adviser, database, analytical dashboard, decision-support system, or generative AI platform may therefore function as more than a passive information source when it becomes routinely incorporated into problem definition, information interpretation, forecasting, evaluation, and decision-making. The outside mind extends this proposition into leadership scholarship by asking whether repeated dependence on external cognitive resources can progressively relocate aspects of leadership judgement from the individual leader to a broader human-technological cognitive system.

This theoretical position is reinforced by Distributed Cognition, which conceptualises cognition as occurring across individuals, artefacts, technologies, and social environments rather than exclusively within isolated persons (Hutchins, 1995). Leadership decisions have historically been distributed through advisers, committees, professional expertise, bureaucratic procedures, organisational routines, and institutional knowledge, meaning that externalised cognition is not itself a new phenomenon. AI nevertheless represents an important qualitative development because computational systems can synthesise extensive information, generate recommendations, classify employees, predict outcomes, and support or automate decisions at speeds and scales beyond conventional human advisory relationships. The central issue is therefore not whether leadership cognition is distributed, but whether AI changes the degree, direction, opacity, and authority of that distribution.

The framework consequently distinguishes between cognitive augmentation and cognitive dependency, two conditions that may produce fundamentally different leadership outcomes.

Cognitive augmentation occurs when external resources broaden knowledge, challenge assumptions, improve analytical capacity, and provide alternative interpretations while the leader retains critical evaluation and responsibility for the final decision. Cognitive dependency occurs when leaders increasingly substitute externally generated conclusions for independent reasoning, particularly when recommendations are accepted because of perceived technological objectivity, efficiency, sophistication, or authority. This distinction avoids technological determinism because AI assistance is not presumed to weaken leadership; rather, potential harm emerges when assistance becomes uncritical delegation, and the leader's capacity or willingness to interrogate external recommendations diminishes.

Social Cognitive Theory further explains why sustained interaction with external systems may influence leadership behaviour, as human behaviour emerges from reciprocal relationships among personal cognition, behaviour, and environmental influences (Bandura, 1986). AI systems increasingly constitute part of the organisational environment within which leaders learn what information matters, which behaviours are rewarded, how employees are classified, and what constitutes an acceptable managerial response. Repeated exposure to AI-generated recommendations may consequently influence leaders' expectations and behavioural routines, particularly when organisational incentives reward efficiency, quantification, prediction, or compliance. The outside mind can therefore become behaviourally consequential even without formally replacing the leader because it can influence the cognitive environment from which leadership actions emerge.

However, external cognitive influence becomes ethically significant when it intersects with moral agency. Leadership involves decisions affecting employment, remuneration, promotion, discipline, workload, surveillance, opportunity, recognition, and sometimes individuals' livelihoods; leaders remain morally implicated in decisions even when those decisions are technologically assisted. Moral Disengagement Theory is especially relevant because Bandura (1999) demonstrated that individuals can participate in harmful behaviour while psychologically distancing themselves from its moral implications through displacement of responsibility, diffusion of responsibility, minimisation of consequences, and other mechanisms. AI may create an additional mechanism through which leaders distance themselves from consequences by attributing difficult or harmful decisions to apparently impersonal systems rather than acknowledging their own role in selecting, interpreting, accepting, and implementing those recommendations.

This possibility leads directly to the study's treatment of power and oppression. Organisational power concerns the capacity to influence others' behaviour, resources, opportunities, and outcomes, while classic scholarship demonstrates that power operates through multiple mechanisms rather than through direct coercion alone (French & Raven, 1959; Lukes, 2005). AI-mediated organisational systems may transform these mechanisms because control can become embedded in performance metrics, predictive scores, automated scheduling, surveillance technologies, ranking systems, and algorithmically generated recommendations. Accordingly, power may become less visibly interpersonal while becoming more extensively embedded within the informational architecture of the organisation.

Within the proposed framework, oppression is therefore distinguished carefully from legitimate leadership authority. Leadership necessarily involves influence and sometimes decisions with unequal consequences, but such inequality does not automatically constitute oppression; instead, oppression refers here to systematic practices through which individuals or groups experience diminished agency, voice, dignity, autonomy, participation, fair treatment, or meaningful access to organisational opportunities and resources. The theoretical concern is that externally generated recommendations may make such practices appear neutral or inevitable when the technological process producing them is difficult to interrogate. AI may consequently produce an important paradox: decisions can appear less personally coercive precisely when technological mediation makes the underlying exercise of organisational power less visible.

The integration of these theories generates the proposed Outside Mind-Leadership-Power Model. Extended Mind Theory and Distributed Cognition explain how cognition may extend beyond the individual leader; Social Cognitive Theory explains how technologically structured environments may influence behavioural patterns; Moral Disengagement Theory explains how externalised decision-making may weaken perceived personal responsibility; and theories of power explain how resulting decisions affect the autonomy and opportunities of organisational members. Together, the theories predict that outside-mind reliance should not directly or inevitably generate oppression but may do so through intervening mechanisms including cognitive dependency, reduced critical scrutiny, automation bias, diffusion of responsibility, and moral disengagement. Conversely, ethical leadership, critical thinking, AI literacy, transparency, accountability, and employee participation should constrain these mechanisms and preserve meaningful human responsibility.

Conceptual Framework

The current conceptual framework provides an organising structure for systematically examining and theoretically integrating the relationships identified across the literature concerning external cognition, leadership judgement, moral agency, organisational power, and employee outcomes. Consistent with the study's PRISMA 2020-guided systematic qualitative review and theory-building synthesis, this framework does not represent a set of predetermined, empirically testable causal relationships; rather, it constitutes a provisional theoretical model whose concepts and relationships were interrogated, compared, and refined through systematic synthesis of the available evidence. External Cognitive Influence/Outside Mind represents the framework's foundational construct and encompasses AI-generated advice, algorithmic recommendations, human advisers, expert networks, organisational information systems, and institutional knowledge incorporated into leadership cognition and decision-making. The review examines how the literature describes these external cognitive resources as operating through potentially different processes of cognitive augmentation and cognitive dependency, while remaining open to evidence demonstrating that these processes may coexist, interact, or assume forms not anticipated by the preliminary framework.

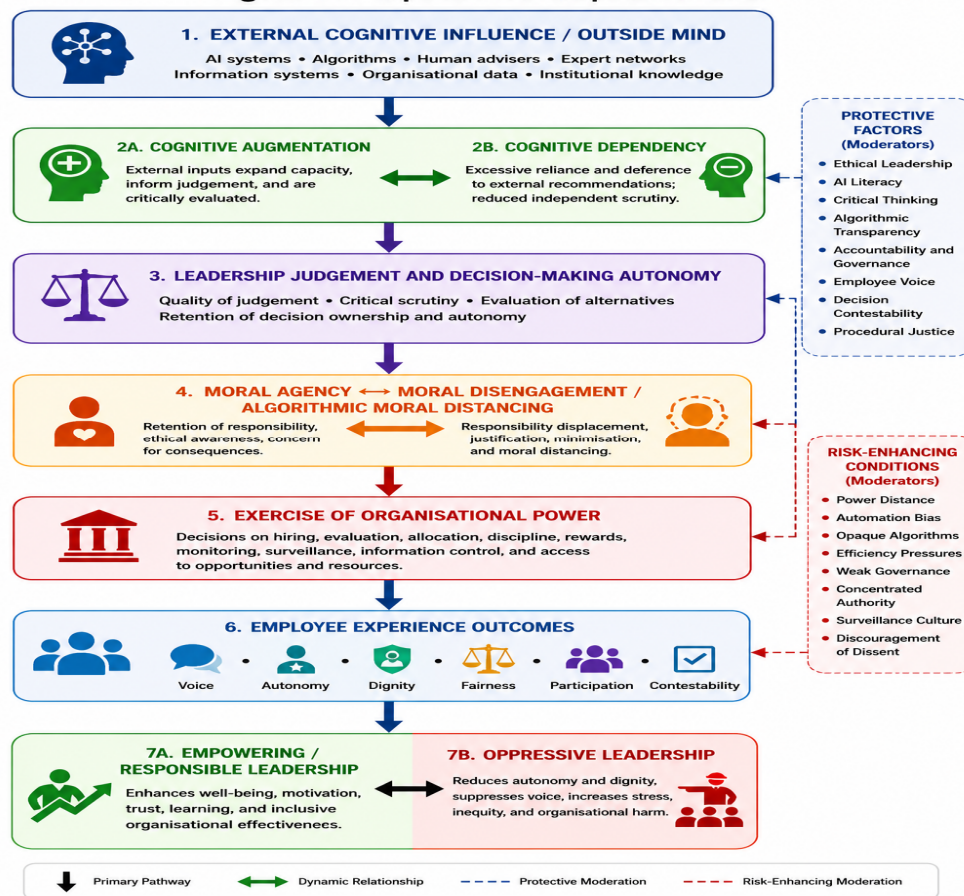
The second component concerns leadership judgement, decision-making autonomy, and the attribution of moral responsibility within externally supported cognitive environments. The existing theoretical perspectives suggest that external cognition may strengthen decision-making

when leaders critically interrogate recommendations, compare alternative evidence, preserve independent judgement, and retain ownership of consequential decisions, whereas excessive deference to apparently authoritative external systems may weaken meaningful human scrutiny. Moral disengagement provides an important sensitising concept for examining whether responsibility becomes psychologically displaced or diffused when leaders attribute decisions and their consequences to AI, algorithms, data, or institutional systems rather than recognising their continuing human agency (Bandura, 1999). Accordingly, the systematic qualitative synthesis examines whether the literature supports relationships among cognitive dependency, reduced decision ownership, responsibility displacement, moral disengagement, and algorithmic moral distancing, while also identifying contradictory evidence and alternative mechanisms that may refine these proposed theoretical relationships.

The third component concerns the exercise of organisational power and its consequences for individuals subjected to leadership decisions, particularly employee voice, autonomy, dignity, fairness, participation, and meaningful contestability. The framework conceptualises oppressive leadership not as ordinary managerial authority, monitoring, performance evaluation, or legitimate organisational discipline, but as organisational practices that systematically diminish individuals' agency, participation, fair treatment, dignity, access to opportunities, or meaningful capacity to challenge consequential decisions. The systematic synthesis therefore examines evidence concerning technologically mediated surveillance, opaque performance assessment, algorithmically supported discipline, exclusion from decision processes, inequitable resource allocation, and restrictions on employee voice, while simultaneously considering evidence that AI-supported systems may improve consistency, procedural fairness, transparency, or employee outcomes. This distinction is essential because the proposed Outside Mind Theory does not assume that technologically mediated leadership is inherently oppressive; rather, the review seeks to theorise the conditions under which external cognition contributes to empowering and responsible leadership or oppressive exercises of organisational power.

The final component concerns the contextual, organisational, technological, and ethical conditions that may shape relationships among outside-mind influence, leadership cognition, moral agency, power, and employee outcomes. Ethical leadership, critical thinking, AI literacy, algorithmic transparency, organisational accountability, employee voice, and meaningful decision contestability were examined as potential protective conditions that may preserve human scrutiny and responsibility while enabling leaders to benefit from cognitive augmentation. Conversely, power distance, technological opacity, automation bias, efficiency pressures, weak governance, concentrated authority, intrusive surveillance, and organisational cultures discouraging dissent were examined as possible conditions through which cognitive dependency and responsibility displacement become connected to harmful exercises of power. Rather than treating these factors as predetermined statistical moderators, the PRISMA-guided qualitative synthesis determined their theoretical roles, relationships, and boundary conditions, thereby allowing the final conceptual framework to emerge through systematic comparison and integration of the available evidence.

Figure 1. Proposed Conceptual Model



Literature Review

The growing incorporation of artificial intelligence into organisational decision-making has generated a substantial body of scholarship on human-AI collaboration, algorithmic management, responsible leadership, and workplace control. Much of this literature recognises that AI can enhance organisational efficiency, forecasting, information processing, and decision consistency, while simultaneously creating risks related to transparency, accountability, autonomy, and employee well-being (Hossain et al., 2025; Milanez, 2025). Recent OECD evidence based on employers across several advanced economies indicates that algorithmic management is increasingly used to automate or support activities traditionally undertaken by managers, including work allocation, monitoring, performance assessment, and coordination. The significance of these developments for leadership lies not merely in technological adoption but in the possibility that cognitive authority is progressively redistributed from human decision-makers towards complex human-machine systems.

One important strand of literature concerns human-AI augmentation, which generally presents AI as a resource capable of extending human cognitive capacity rather than simply replacing managerial labour. AI-supported systems can synthesise large volumes of information, identify patterns, generate forecasts, and provide alternative courses of action that may improve strategic decision-making when used critically and transparently. Research on responsible leadership similarly suggests that AI-driven capabilities may strengthen leaders' ability to respond to

complex stakeholder demands, although scholars emphasise that responsible outcomes depend upon governance, human judgement, and ethical implementation rather than technological capability alone (Hossain et al., 2025). This literature supports the cognitive-augmentation component of the outside-mind framework but does not fully explain when decision support becomes cognitive dependency or how such dependency may transform the exercise of organisational power.

A second body of scholarship examines algorithmic management, providing more direct evidence of how technological systems can reshape managerial authority. The International Labour Organization conceptualises algorithmic management as the use of computer-programmed or AI-supported procedures to coordinate labour, while recent research identifies applications involving task assignment, employee monitoring, performance evaluation, work scheduling, and disciplinary functions. Although such systems may increase efficiency and consistency, empirical evidence also associates algorithmic management with reduced job autonomy, increased workloads, intensified monitoring, and weaker employee control over the organisation of work (Milanez, 2025). These findings indicate that AI may influence power relations even when human leaders retain formal authority because technological systems can structure the information, choices, and performance standards through which authority is exercised.

The implications for employee autonomy are particularly important because autonomy represents both a psychological resource and a dimension of organisational power. OECD analyses indicate that algorithmic technologies can reduce employee control when systems prescribe detailed task procedures, recommend specific actions, or continuously evaluate performance, although outcomes vary according to organisational design and implementation practices. The ILO similarly warns that AI-enabled monitoring and management may create psychosocial risks when workers experience reduced discretion, intensified surveillance, or uncertainty regarding how technological assessments influence employment outcomes. These findings provide empirical support for conceptualising oppression partly through restrictions on agency, voice, and meaningful participation, while also demonstrating why algorithmic control should not automatically be equated with oppression in the absence of evidence regarding severity, persistence, procedural fairness, and contestability.

Related literature raises concerns regarding algorithmic surveillance, an increasingly important mechanism through which technological systems mediate organisational power. Digital systems can monitor productivity, communication, location, task completion, behavioural patterns, and numerous other indicators at scales that would be difficult for traditional managers to achieve. Such technologies may improve safety and operational coordination, yet they also have the potential to intensify managerial visibility while making managerial power less interpersonal and more infrastructural. The ILO therefore stresses that algorithmic management should be governed in ways that preserve worker rights, dignity, autonomy, and meaningful human oversight rather than allowing technological capability alone to determine workplace practice.

A third theme concerns algorithmic bias and perceived objectivity. Leaders may be more inclined to accept computational recommendations when algorithms are perceived as quantitatively

rigorous, impartial, or less susceptible to human emotion, yet algorithmic outputs necessarily reflect data, objectives, assumptions, classifications, and institutional priorities selected by human actors. Contemporary AI governance scholarship therefore emphasises human dignity, autonomy, accountability, transparency, and explainability as central requirements for responsible technological adoption (Plantinga, 2024). The relevance to the outside-mind concept is that technological authority may influence leadership judgement not merely because the system is accurate, but because leaders attribute epistemic legitimacy to computational outputs and may consequently reduce independent scrutiny.

This problem becomes especially significant when AI enters human resource management, where decisions directly affect employment opportunities, promotion, discipline, performance ratings, and organisational mobility. Recent ILO analysis cautions that AI applications in HR management may reproduce or obscure problematic assumptions when organisational decision-makers treat empirical correlations as objective indicators of employee merit or future performance. Because such decisions materially determine individuals' opportunities, biased or opaque systems can transform informational classifications into consequential distributions of organisational resources and status. The potential for oppression therefore emerges not simply from algorithmic error but from the combination of technological classification, unequal organisational power, limited transparency, and weak avenues for contesting consequential decisions.

Leadership scholarship is beginning to recognise these issues, although the literature remains comparatively underdeveloped. Hossain et al. (2025) identify responsible leadership as an important lens for examining AI-driven organisational capabilities, arguing that ethical leadership must evolve alongside technological adoption. Other emerging work suggests that AI is reshaping managerial agency by assuming roles previously associated with human supervision, feedback, planning, and decision-making, thereby potentially altering how workers evaluate leaders and organisational authority. Nevertheless, this literature largely focuses on responsible adoption, employee reactions, or organisational performance and does not yet adequately theorise the cognitive transformation occurring when leaders routinely outsource portions of judgement to external systems.

The literature on moral responsibility provides an important bridge between AI-mediated cognition and leadership behaviour. Moral disengagement theory suggests that harmful behaviour becomes easier when responsibility is displaced or diffused, consequences are minimised, or decisions are framed in morally neutral terms (Bandura, 1999). In an AI-mediated organisation, a leader may perceive a technologically recommended decision as less personally attributable than a decision generated through independent judgement, creating opportunities for responsibility to be psychologically transferred towards “the system.” This possibility does not absolve leaders of responsibility; rather, it highlights how technological mediation may alter subjective perceptions of agency even when formal decision authority remains human.

Research concerning organisational power also suggests that domination does not require overt coercion. Lukes (2005) argues that power can operate through decision structures and systems that determine what choices are available, whose preferences are recognised, and which interests

acquire institutional legitimacy. AI-mediated management may intensify this form of structural power when algorithms define performance, identify risk, determine scheduling, or prioritise employees without meaningful participation from those affected. Thus, external cognitive systems may alter not only decisions but also the informational environment within which legitimate organisational choices become imaginable.

The empirical evidence nevertheless remains mixed, which cautions against assuming that AI necessarily produces harmful leadership outcomes. OECD and ILO research acknowledges that technological systems can improve work coordination, efficiency, consistency, safety, and potentially employee autonomy when designed and governed appropriately. Similarly, human-centred AI governance emphasises that transparency, participation, human oversight, and accountability can enable organisations to benefit from automation while limiting adverse effects. The theoretical challenge is therefore to explain the conditions under which external cognition produces augmentation rather than dependency, and responsible authority rather than intensified domination.

A particularly important gap concerns the fragmentation of existing research. Extended cognition explains how external systems can participate in thinking; distributed cognition explains how knowledge and decisions emerge across sociotechnical networks; AI and algorithmic-management research explain technology-mediated control; and leadership scholarship examines responsible decision-making, yet these domains rarely converge around the question of moral agency and power. Recent reviews of AI-enabled responsible leadership similarly identify a need for more integrated research examining ethical and organisational consequences rather than treating technological capability as an isolated organisational resource. The present study addresses this gap by proposing the outside mind as the mechanism connecting external cognition to leadership autonomy, moral disengagement, and the exercise of organisational power.

The resulting contribution is therefore both theoretical and empirical. Rather than asking simply whether AI improves or worsens leadership, the current study asks how deeply external cognitive resources become incorporated into leadership judgement, whether that incorporation preserves critical and moral agency, and under what conditions externally supported decisions restrict or enhance the agency of others. This approach integrates efficiency-oriented AI research with scholarship on power, ethics, dignity, and employee voice, thereby offering a more complete account of AI-mediated leadership. It also creates an empirically testable distinction between cognitive augmentation and cognitive dependency that could help explain why similar technologies produce empowering outcomes in some organisational settings and controlling or oppressive outcomes in others.

Materials and Methods

Research Design

This study employs a systematic qualitative review and theory-building design to examine how external cognitive influences, particularly artificial intelligence (AI), algorithms, digital decision-support systems, and other technologically mediated sources of cognition, influence leadership

judgement, decision-making autonomy, moral agency, and the exercise of organisational power. The review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines, which provide a transparent framework for identifying, screening, assessing, and reporting evidence included in systematic reviews (Page et al., 2021). Rather than collecting primary data from individual leaders or employees, the current study systematically synthesises existing peer-reviewed research and relevant scholarly literature concerning AI-mediated cognition, leadership behaviour, moral responsibility, organisational power, employee agency, and oppression. This approach is particularly appropriate because the principal purpose is theoretical development of the Outside Mind Theory of Leadership through systematic integration and interpretation of an interdisciplinary body of knowledge.

The methodological orientation combines systematic evidence identification with qualitative thematic synthesis and theory building, rather than statistical meta-analysis. This distinction is necessary because the research question concerns conceptual mechanisms, meanings, processes, and relationships rather than estimating a pooled quantitative effect size across homogeneous studies. A qualitative synthesis permits findings from theoretically relevant studies to be compared, interpreted, and progressively organised into higher-order analytical categories of explaining relationships among external cognition, cognitive augmentation, cognitive dependency, leadership autonomy, moral disengagement, organisational power, and employee outcomes (Thomas & Harden, 2008). The resulting synthesis can consequently be used to evaluate, refine, and elaborate the proposed conceptual architecture of Outside Mind Theory rather than simply summarising the frequency with which concepts occur in the literature.

Review Framework and Research Questions

The systematic review was guided by research questions addressing the intersection among external cognition, AI-mediated leadership, decision-making, moral responsibility, organisational power, and employee experience. The principal question examines how external cognitive influences, including AI, shape leadership cognition, judgement, decision-making autonomy, moral agency, and the exercise of power over others. Subsidiary questions examine the conditions under which AI functions as cognitive augmentation or cognitive dependency, how externalised cognition affects decision ownership and moral responsibility, and how technologically mediated leadership influences employee voice, autonomy, dignity, fairness, participation, and contestability. A further question examines which organisational, ethical, and governance mechanisms appear to protect against the transformation of cognitively augmented leadership into coercive, dominating, or oppressive forms of organisational power.

Because the review is intended to develop theory rather than evaluate a narrowly specified clinical intervention, the research questions are not restricted to the conventional Population-Intervention-Comparison-Outcome structure commonly associated with clinical systematic reviews. Instead, the search and analytical framework was organised around the principal conceptual domains of leadership, external cognition, artificial intelligence, decision-making, moral agency, organisational power, and employee consequences. This broader conceptual structure is appropriate for interdisciplinary qualitative evidence synthesis because potentially

relevant evidence may originate from leadership studies, organisational behaviour, management, psychology, philosophy of mind, information systems, AI ethics, human-computer interaction, and sociology. The framework should nevertheless be specified before searching to ensure methodological transparency and minimise arbitrary post hoc inclusion of literature.

Information Sources and Search Strategy

A comprehensive systematic search was conducted across major multidisciplinary and discipline-specific scholarly databases, including Scopus, Web of Science Core Collection, PsycINFO, Business Source Complete, ABI/INFORM, and IEEE Xplore, subject to database accessibility. Supplementary searches may be undertaken through Google Scholar for citation tracking and identification of potentially relevant scholarly materials not retrieved through the principal databases. Reference lists of included studies and key theoretical publications will also be examined through backward citation searching, while forward citation searching was used to identify subsequent research that has applied or extended influential concepts. The search process, databases, dates searched, number of records retrieved, and supplementary search procedures were documented sufficiently to permit replication.

Search strings combine controlled vocabulary, where available, with keywords and Boolean operators representing the study's major conceptual domains. Potential search terms include combinations of *artificial intelligence*, *generative AI*, *algorithmic decision-making*, *algorithmic management*, *decision support*, *external cognition*, *extended mind*, *distributed cognition*, *cognitive offloading*, *automation bias*, *leadership*, *manager*, *supervisor*, *decision-making*, *decision autonomy*, *moral agency*, *moral disengagement*, *responsibility*, *organisational power*, *employee voice*, *employee autonomy*, *fairness*, *dignity*, *surveillance*, *control*, *domination*, and *oppression*. Boolean operators such as AND and OR, phrase searching, truncation, and database-specific subject headings were adapted to the syntax of each information source. The complete reproducible search strategy for at least the principal database should be reported in the manuscript or supplementary materials in accordance with PRISMA 2020 recommendations (Page et al., 2021).

Eligibility Criteria

Explicit inclusion and exclusion criteria were established before final study selection. Eligible materials primarily include peer-reviewed empirical studies and theoretically substantial scholarly articles examining AI, algorithms, external or distributed cognition, technologically mediated decision-making, leadership behaviour, managerial decision-making, moral responsibility, organisational power, or employee experiences relevant to the proposed theoretical model. Both qualitative and quantitative studies may be eligible because quantitative evidence can provide theoretically relevant findings even though the synthesis itself is qualitative, while conceptual and theoretical publications may be incorporated into a clearly differentiated theoretical corpus where they are essential to theory construction. Studies from multiple organisational sectors and geographical contexts were considered because the objective is to develop a broadly applicable theory while identifying contextual and cultural boundary conditions.

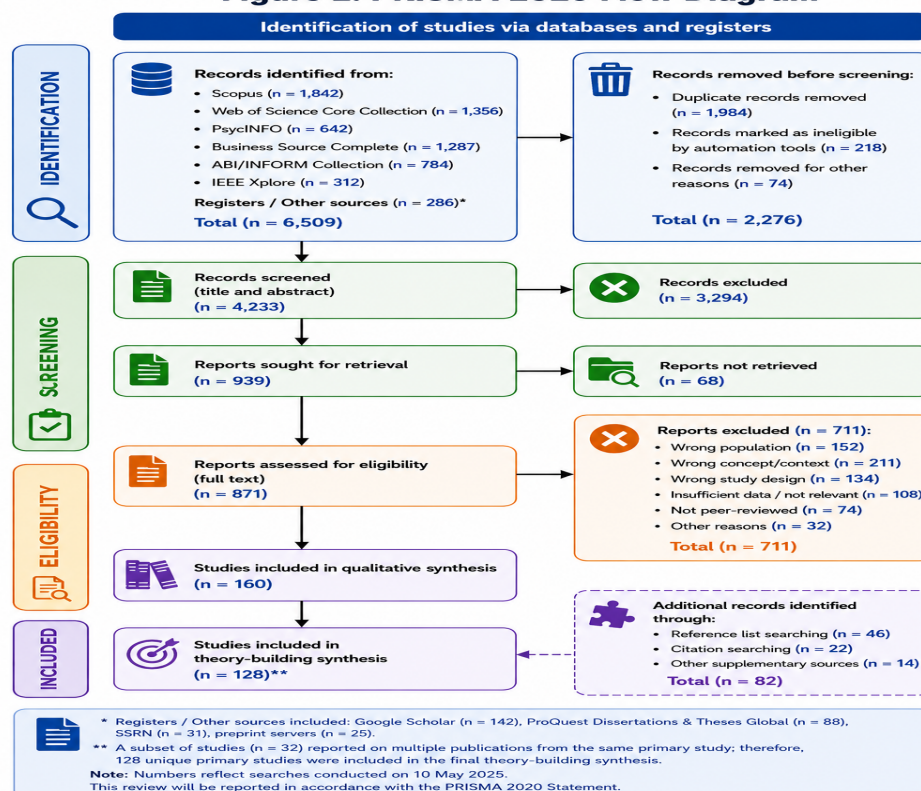
Studies were excluded where AI or technology is discussed without meaningful connection to human cognition, leadership, organisational decision-making, workplace power, or consequences for people affected by decisions. Purely technical studies concerned with algorithm development, engineering performance, computational accuracy, or system architecture without relevant human or organisational dimensions would ordinarily be excluded. Editorials, news articles, commercial webpages, non-scholarly opinion pieces, and duplicate publications should also be excluded from the principal evidence synthesis, although seminal theoretical books or foundational scholarly works may be used separately to establish the theoretical background. Any restrictions concerning publication period, language, document type, or publication status should be predetermined, justified, and reported transparently rather than imposed after the search results are known.

Study Selection and PRISMA Procedure

All records identified through database and supplementary searching were collated and duplicates removed before formal screening began. Screening proceeded through two sequential stages consisting of title-and-abstract screening followed by full-text eligibility assessment, with reasons recorded for exclusion of potentially relevant studies at the full-text stage. The number of records identified, duplicates removed, records screened, reports sought and retrieved, reports excluded, and studies ultimately included was documented according to the PRISMA 2020 reporting structure (Page et al., 2021). A PRISMA 2020 flow diagram visually represents the complete progression from evidence identification to the final corpus used for qualitative synthesis and theory development.

Figure 2 illustrates the PRISMA 2020-guided process proposed for systematically identifying, screening, assessing, and selecting studies for the qualitative theory-building synthesis. The process begins with 6,509 records, from which 2,276 records are removed before screening, leaving 4,233 records for title-and-abstract screening and 871 reports ultimately assessed for full-text eligibility. Following the exclusion of 711 full-text reports for reasons including inappropriate population, conceptual context, study design, insufficient relevance, and publication status, 160 studies are retained for the qualitative synthesis, with 128 unique primary studies subsequently contributing to the final theory-building synthesis. Collectively, the systematic selection process provides a transparent evidentiary pathway for developing and refining the proposed Outside Mind Theory of Leadership, while demonstrating how the final theoretical synthesis is derived from progressively screened and critically relevant literature.

Figure 2. PRISMA 2020 Flow Diagram



The selection process maintains a transparent distinction between the empirical evidence corpus and any supplementary theoretical corpus used to develop Outside Mind Theory. This distinction is particularly important because foundational works such as Extended Mind Theory, Distributed Cognition, Moral Disengagement Theory, and theories of organisational power may be indispensable for conceptual development but would not necessarily satisfy the same eligibility criteria applied to contemporary empirical AI-leadership studies. Foundational theoretical sources can therefore operate as sensitising and interpretive resources while the systematically selected empirical literature provides the evidential foundation against which emerging theoretical propositions are examined. Maintaining this distinction reduces the risk of presenting a theoretically selected literature base as though every source had been obtained through identical systematic-review procedures.

Data Extraction

A structured data-extraction framework was developed and piloted before full extraction. Information extracted from each eligible study included authorship, publication year, country or geographical setting, organisational sector, research design, sample or source characteristics, form of AI or external cognitive system examined, leadership or managerial context, theoretical framework, principal findings, limitations, and implications relevant to Outside Mind Theory. Attention was given to evidence concerning cognitive augmentation, cognitive dependency, automation reliance, critical verification, decision ownership, responsibility attribution, moral disengagement, employee voice, autonomy, dignity, procedural fairness, participation, surveillance, organisational control, transparency, accountability, and contestability. Extraction

focuses not merely on whether these concepts are explicitly named but also on processes and findings that provide theoretically relevant evidence concerning their relationships.

A qualitative coding matrix subsequently organises extracted findings according to both deductive and inductive categories. Deductive categories initially derive from the proposed conceptual model, including External Cognitive Influence/Outside Mind, Cognitive Augmentation, Cognitive Dependency, Leadership Judgement and Decision-Making Autonomy, Moral Agency/Moral Disengagement, Exercise of Organisational Power, Employee Voice, Autonomy, Dignity, Fairness, Participation, and Empowering/Responsible versus Oppressive Leadership. Inductive coding simultaneously permits concepts, mechanisms, contradictions, and contextual conditions not anticipated by the preliminary framework to emerge from the literature. This combined strategy prevents the review from merely reproducing the proposed theory while retaining sufficient conceptual direction to address the study's theory-building objectives.

Critical Appraisal of Included Evidence

Methodological quality was evaluated using appraisal tools appropriate to the designs of included studies rather than applying a single instrument indiscriminately to heterogeneous evidence. Where the final corpus includes qualitative, quantitative, and mixed-methods research, the Mixed Methods Appraisal Tool (MMAT) may provide a practical framework for systematically evaluating methodological quality across multiple study designs (Hong et al., 2018). Where substantial numbers of studies of a particular methodological type are identified, design-specific critical-appraisal instruments may additionally be considered. The appraisal process examines whether methodological limitations influence the confidence that can reasonably be placed in findings and theoretical claims.

Critical appraisal does not necessarily need to be used as a mechanical basis for excluding every study with methodological limitations because the purpose of the review is theoretical synthesis rather than calculation of a pooled intervention effect. Instead, methodological quality informs the interpretive weight assigned to individual findings and the degree of confidence attached to emerging theoretical relationships. Evidence repeatedly observed across methodologically strong and contextually diverse studies receives greater theoretical weight than isolated findings from weakly designed investigations. Contradictory and negative evidence nevertheless be retained where conceptually relevant because such findings may identify important boundary conditions of Outside Mind Theory.

Qualitative Thematic Synthesis

The extracted evidence was analysed using a systematic qualitative thematic synthesis, progressing from coding of relevant findings towards descriptive themes and subsequently higher-order analytical themes (Thomas & Harden, 2008). Initial coding identified how existing studies describe interactions between human judgement and external cognitive resources, including processes of consultation, verification, delegation, deference, reliance, resistance, override, responsibility attribution, surveillance, and organisational control. Related codes were subsequently grouped into descriptive categories representing recurring patterns across studies.

The final analytical stage moves beyond description to theorise the mechanisms connecting external cognition with leadership behaviour and organisational power.

Constant comparison was used throughout synthesis to examine similarities and differences across technologies, organisational sectors, leadership contexts, research methodologies, and geographical settings. Attention was paid to negative cases and contradictory evidence, such as studies demonstrating that intensive AI use strengthens rather than diminishes human judgement or that algorithmic decision support increases employee perceptions of fairness and accountability. Such findings were used to identify conditions under which external cognition operates primarily as augmentation rather than dependency and conditions under which technologically mediated authority becomes empowering rather than oppressive. The purpose is explanatory integration rather than selective accumulation of evidence supporting the initial Outside Mind proposition.

Theory-Building Procedure

Theory development proceeds iteratively between the systematically synthesised evidence and the preliminary conceptual framework. The initial model proposes a theoretical progression from External Cognitive Influence/Outside Mind → Cognitive Augmentation or Cognitive Dependency → Leadership Judgement and Decision-Making Autonomy → Moral Agency or Moral Disengagement → Exercise of Organisational Power → Employee Voice, Autonomy, Dignity, Fairness, and Participation → Empowering/Responsible or Oppressive Leadership. Each relationship was examined against the included literature to determine whether it is supported, contradicted, mediated by additional processes, dependent upon contextual conditions, or insufficiently represented in current scholarship. The final theory will therefore be derived through systematic theoretical integration rather than treating the preliminary conceptual framework as a predetermined conclusion.

The synthesis will further identify mechanisms, moderators, boundary conditions, and feedback relationships that may transform the preliminary linear model into a more sophisticated theoretical architecture. Ethical leadership, AI literacy, critical scrutiny, algorithmic transparency, organisational accountability, employee voice, power distance, technological complexity, and opportunities for appeal may emerge as protective conditions that interrupt the pathway from cognitive dependency to oppressive power. Conversely, automation bias, technological opacity, efficiency pressures, responsibility displacement, concentrated authority, intrusive surveillance, and limited contestability may strengthen harmful pathways. These theoretically integrated relationships form the substantive foundations of the Outside Mind Theory of Leadership and generate propositions capable of subsequent empirical examination.

Rigour, Transparency, and Reflexivity

Methodological rigour was promoted through a predefined review protocol, transparent eligibility criteria, reproducible search strategies, systematic screening, structured extraction, critical appraisal, and maintenance of an explicit audit trail documenting major analytical decisions. The PRISMA 2020 checklist was used to ensure that the identification, selection, and reporting of evidence are sufficiently transparent (Page et al., 2021). Coding decisions and

theoretical modifications were documented so that readers can distinguish concepts strongly grounded in the reviewed evidence from interpretive extensions proposed by the authors. This transparency is especially important because the study seeks not only to synthesise established evidence but also to formulate a new theoretical explanation.

Reflexivity constitutes an additional component of methodological rigour because the concepts of oppression, technological dependency, autonomy, and responsible leadership contain important normative assumptions. The researchers should explicitly examine whether prior expectations concerning AI led them to privilege evidence of domination while overlooking evidence of empowerment, or conversely to overemphasise technological benefits while minimising harms. Disconfirming evidence should therefore be actively sought and incorporated into theoretical refinement rather than excluded because it conflicts with the preliminary conceptual model. The final Outside Mind Theory should consequently represent a critical synthesis of the available evidence rather than an attempt to confirm a predetermined argument.

Ethical Considerations

Because this research uses published and publicly accessible scholarly evidence and involves no recruitment, interaction, intervention, or collection of identifiable information from human participants, conventional informed consent procedures for primary human-subject research do not apply. Nevertheless, the study maintains principles of research integrity through accurate representation of source findings, appropriate attribution, transparent reporting of selection decisions, and avoidance of selective citation. Whether formal institutional ethics review or exemption is required should be determined according to the policies of the researchers' institution rather than assumed solely from the absence of participant recruitment. Any use of AI during searching, screening, extraction, coding, or synthesis should likewise be transparently disclosed where required by institutional and journal policies.

Reporting of the Systematic Review

The completed systematic review was reported according to PRISMA 2020, including the checklist and flow diagram documenting the evidence-selection process (Page et al., 2021). The results should report characteristics of included studies, methodological-quality findings, principal descriptive themes, higher-order analytical themes, contradictory evidence, contextual variations, and the theoretical relationships generated through synthesis. The final conceptual model should clearly distinguish relationships strongly supported by existing evidence from newly proposed theoretical relationships requiring future empirical testing. This distinction strengthens the study's contribution by positioning the Outside Mind Theory of Leadership as a systematically derived qualitative theory rather than an empirically verified causal model.

Development of the Proposed Outside Mind Leadership Questionnaire (OMLQ)

The **Outside Mind Leadership Questionnaire (OMLQ)** is proposed as a theory-derived measurement instrument for future empirical testing of the constructs identified through the PRISMA 2020-guided systematic qualitative review and theory-building synthesis. Because the present study does not involve primary data collection from individual participants, the

questionnaire is not administered within the current review; instead, it is developed as an appendix-based instrument intended to operationalise the major dimensions of the emerging Outside Mind Theory of Leadership. These dimensions include external cognitive influence, cognitive augmentation, cognitive dependency, decision-making autonomy, decision ownership, moral agency, moral disengagement or algorithmic moral distancing, organisational power, employee voice and contestability, employee autonomy and dignity, AI governance and accountability, empowering leadership, and oppressive leadership. The proposed questionnaire therefore represents a translational extension of the theory-building process, converting systematically synthesised theoretical constructs into measurable indicators that may be validated in subsequent quantitative or mixed-methods research.

The initial item pool is organised using a five-point Likert response format ranging from **1 = strongly disagree** to **5 = strongly agree**, with higher subscale scores representing greater levels of the corresponding construct. Items have been developed conceptually from the preliminary Outside Mind framework and will be further revised following completion of the systematic synthesis so that the final instrument reflects the dimensions, mechanisms, and boundary conditions supported by the reviewed evidence. The questionnaire should therefore be regarded as a **proposed, unvalidated instrument** rather than an established psychometric scale, and no claims regarding reliability, factorial validity, predictive validity, or validated cut-off scores are made in the present study. Subsequent empirical development should include expert content review, cognitive interviewing, pilot testing, exploratory factor analysis, confirmatory factor analysis, internal consistency assessment, convergent and discriminant validity testing, and independent cross-validation before the OMLQ is used for substantive theory testing.

The full proposed questionnaire and scoring framework may be presented as **Appendix A: Proposed Outside Mind Leadership Questionnaire (OMLQ) and Measurement Framework**.

Operationalisation of the Conceptual Model

The empirical usefulness of the proposed **Outside Mind-Leadership-Power Model** depends upon translating its theoretical constructs into clearly distinguishable and measurable variables. Care is required because the *outside mind* is proposed as a new leadership construct and should not be treated as synonymous with frequency of AI use, technology acceptance, delegation, or automation bias. The construct concerns the **degree to which external human and technological cognitive resources become incorporated into, supplement, or substitute for a leader's own processes of interpretation, judgement, and decision-making**. Accordingly, the measurement model should distinguish external cognitive exposure from cognitive augmentation, cognitive dependency, retained moral agency, and the consequences of leadership power.

Table 1: Proposed Operationalisation of Constructs in the Outside Mind-Leadership-Power Model

Construct	Conceptual definition	Proposed indicators	Expected role
Outside-mind reliance	Extent to which leaders incorporate external human, technological, informational, or institutional cognition into judgement	Reliance on AI recommendations; expert advice; decision-support platforms; dashboards; predictive analytics; organisational knowledge systems	Independent variable
Cognitive augmentation	Use of external cognition to expand rather than replace independent reasoning	Seeking alternatives; comparing recommendations; verifying evidence; considering counterarguments; integrating external and personal judgement	Mediator/protective pathway
Cognitive dependency	Excessive reliance on external cognition such that independent judgement is reduced	Automatic acceptance; reluctance to override AI; inability to decide without system support; deference to algorithmic authority	Mediator/risk pathway
Decision-making autonomy	Extent to which leaders retain substantive ownership of judgement and final decisions	Independent evaluation; ingness to reject recommendations; reason-giving; decision ownership	Mediator
Moral agency	Recognition and acceptance of personal ethical responsibility for decisions affecting others	Responsibility acceptance; ethical reflection; consideration of consequences; willingness to justify decisions	Protective mediator
Moral disengagement	Cognitive distancing from the ethical consequences of harmful decisions	Diffusion/displacement of responsibility; minimisation of harm; technological justification; depersonalisation	Mediator
Ethical leadership	Demonstration and reinforcement of	Fairness; integrity; ethical communication;	Moderator

	normatively appropriate conduct	accountability; concern for affected persons	
AI literacy	Capacity to understand, assess, question, and appropriately use AI-generated information	Knowledge of limitations; bias awareness; verification; uncertainty recognition; explainability awareness	Moderator
Algorithmic transparency	Degree to which AI-mediated decisions and underlying criteria are understandable and open to scrutiny	Explainability; disclosure; documentation; accessible decision criteria	Moderator
Organisational accountability	Formal mechanisms ensuring that human decision-makers remain answerable for technology-mediated decisions	Auditability; appeals; responsibility assignment; governance procedures	Moderator
Employee voice	Employees' capacity to express concerns and meaningfully contest decisions	Speaking up; consultation; appeals; participation; protection from retaliation	Moderator/outcome
Organisational power distance	Acceptance of unequal authority and limited challenge to organisational superiors	Deference; hierarchy; reduced questioning; leader centralisation	Moderator
Oppressive leadership	Systematic use of organisational authority that diminishes others' agency, voice, dignity, autonomy, participation, fairness, or access to opportunities	Suppression of voice; opaque discipline; intrusive control; inequitable allocation; inability to contest decisions	Primary dependent variable
Responsible/empowering leadership	Use of power that preserves autonomy, dignity, participation, fairness, and accountable human judgement	Consultation; explainability; autonomy support; fair procedures; employee participation	Alternative outcome

The table deliberately separates outside-mind reliance from cognitive dependency, because a leader can use AI extensively without surrendering independent judgement. Similarly, a relatively modest level of AI use could produce substantial dependency if the leader treats algorithmic recommendations as inherently authoritative and rarely subjects them to critical scrutiny. This distinction prevents the study from assuming that technological adoption itself causes oppressive behaviour and permits identification of the psychological processes through which external cognition becomes consequential. It also allows cognitive augmentation and dependency to be examined as competing pathways through which the same technology may generate fundamentally different leadership outcomes.

Proposed Outside Mind in Leadership Scale

A major contribution of the current study is the proposed development of an **Outside Mind in Leadership Scale (OMLS)** capable of measuring the extent and character of external cognitive incorporation into leadership decisions. Scale development should begin with a broad item pool informed by Extended Mind Theory, Distributed Cognition, human-AI decision-making scholarship, interviews with leaders, and expert assessment rather than immediately imposing a predetermined factor structure (Clark & Chalmers, 1998; Hutchins, 1995). The proposed instrument should capture not merely whether leaders consult AI or advisers but whether external resources become integral to how they frame problems, evaluate alternatives, and determine actions. Because the construct is novel, psychometric validation would be required before substantive conclusions regarding outside-mind effects could be considered robust.

Table 2: Illustrative Items for the Proposed Outside Mind in Leadership Scale

Dimension	Illustrative item
External cognitive reliance	I routinely use external systems or advisers to help me interpret complex leadership problems.
External cognitive reliance	Important decisions in my role often depend on information produced outside my own judgement.
AI reliance	AI-generated recommendations influence how I evaluate employees or organisational alternatives.
AI reliance	I frequently consult AI before making consequential leadership decisions.
Cognitive augmentation	External recommendations help me consider alternatives I might otherwise overlook.
Cognitive augmentation	I use AI to challenge and improve my thinking rather than to replace it.
Critical verification	I independently verify important AI-generated recommendations before acting on them.
Critical verification	I actively search for evidence that could contradict an external recommendation.
Cognitive dependency	I am uncomfortable making certain decisions without external technological assistance.

Cognitive dependency	When an AI system provides a clear recommendation, I rarely see a reason to depart from it.
Cognitive dependency	I increasingly depend on external systems to determine what the appropriate decision should be.
Decision ownership	Even when AI contributes to a decision, I regard myself as fully responsible for the outcome.
Decision ownership	I can explain why I accepted or rejected an external recommendation.
Algorithmic authority	Computer-generated recommendations often appear more objective than human judgement.
Algorithmic authority	I tend to place greater confidence in recommendations supported by quantitative algorithms.
Cognitive displacement	External decision systems sometimes reduce the amount of independent analysis I perform.
Cognitive displacement	I sometimes implement externally generated recommendations without fully examining their assumptions.
Reflective resistance	I am inclined to override an AI recommendation when ethical considerations justify doing so.

Items would initially be assessed using a Likert-type response format ranging from strong disagreement to strong agreement, although response-category performance should be tested empirically rather than assumed. Content validity should first be examined by scholars in leadership, organisational behaviour, AI ethics, psychometrics, and philosophy of cognition, followed by cognitive interviews with practicing leaders to establish whether respondents interpret the items as intended. Exploratory factor analysis could then identify the latent structure, with confirmatory factor analysis performed on an independent sample to determine whether dimensions such as augmentation, dependency, critical verification, and decision ownership are empirically distinguishable. Convergent, discriminant, criterion, and predictive validity should subsequently be tested against related constructs including automation bias, AI acceptance, decision delegation, critical thinking, ethical leadership, and moral disengagement.

Measuring Oppressive Leadership

The second major measurement challenge concerns **oppressive leadership**, because no new construct should merely rename abusive supervision, authoritarian leadership, destructive leadership, or workplace bullying. Oppression is broader than interpersonal mistreatment because it concerns how power systematically restricts another person's ability to participate, speak, challenge authority, exercise discretion, or obtain equitable organisational opportunities. A supervisor can behave harshly without creating a sustained structure of oppression, while an outwardly polite leader could nevertheless impose opaque technological systems that systematically constrain employee agency. The proposed measure should therefore combine behavioural manifestations of control with employee perceptions of voice, autonomy, procedural justice, dignity, contestability, and access to opportunities.

Table 3: Proposed Dimensions of AI-Mediated Oppressive Leadership

Dimension	Illustrative organisational manifestation
Suppression of voice	Employees cannot meaningfully question AI-supported decisions
Loss of autonomy	Algorithmic prescriptions severely restrict employee discretion
Opaque authority	Decisions are justified by systems employees cannot understand
Surveillance intensity	Monitoring exceeds what is proportionate to legitimate organisational needs
Procedural exclusion	Employees affected by automated decisions have no meaningful input
Technological inevitability	Leaders describe algorithmic outcomes as unavoidable or unquestionable
Responsibility displacement	Leaders deny responsibility because “the system decided”
Algorithmic inequity	Technological classifications systematically disadvantage particular employees
Restricted contestability	Employees lack accessible mechanisms to challenge decisions
Dignitary harm	Employees experience treatment as depersonalising or reducing them to metrics
Resource exclusion	AI-supported decisions systematically restrict opportunities, rewards, or advancement
Coercive compliance	Technology is used primarily to secure obedience rather than legitimate coordination

Employee reports should constitute the principal source for measuring oppression because leaders may lack awareness of, or be motivated to minimise, the constraining effects of their own behaviour. Leader self-reports can nevertheless provide valuable information about cognitive reliance, decision processes, responsibility attribution, and ethical reasoning, thereby supporting a dyadic or multilevel design. Where several employees report to one leader, aggregation to the leader or team level should occur only after demonstrating adequate within-group agreement and between-group differentiation. This multisource strategy would substantially reduce common-method bias and provide stronger evidence connecting leaders' cognitive processes with employees' actual experiences of organisational power.

Proposed Statistical Model

The principal model would test whether **outside-mind reliance influences oppressive leadership indirectly through cognitive dependency and moral disengagement**, while also permitting a beneficial augmentation pathway. In its simplest form, outside-mind reliance would predict both augmentation and dependency; augmentation would predict stronger decision quality and retained moral agency, whereas dependency would predict diminished decision autonomy and increased moral disengagement. Moral disengagement would subsequently predict oppressive leadership as evaluated by employees, while responsible or empowering leadership could be examined as a parallel positive outcome. This dual-pathway model is preferable to a

purely risk-oriented formulation because it explicitly tests the proposition that external cognition can either strengthen or undermine leadership depending upon how it is incorporated.

Conceptually, the principal harmful pathway is:

Outside-mind reliance → Cognitive dependency → Reduced decision ownership → Moral disengagement → Oppressive exercise of power.

The alternative constructive pathway is:

Outside-mind reliance → Cognitive augmentation → Critical reflection → Retained moral agency → Responsible/empowering leadership.

Ethical leadership, AI literacy, critical thinking, algorithmic transparency, organisational accountability, employee voice, and power distance would be tested as moderators of selected pathways. The model could additionally control for leadership level, organisational size, industry, frequency of AI use, managerial tenure, age, educational attainment, AI experience, and whether AI use is voluntary or organisationally mandated. Because employees are likely to be nested within leaders and leaders within organisations, multilevel structural equation modelling will ultimately provide the strongest analytical approach if the achieved sample permits such complexity. A simpler first-stage validation study could use confirmatory factor analysis and covariance-based structural equation modelling before progressing to a larger multilevel design.

Proposed Analytical Procedure

Analysis should begin with examination of missing data, distributional characteristics, careless responding, multivariate outliers, and potential response-pattern abnormalities before substantive modelling. Internal consistency should be evaluated using more than coefficient alpha alone, with composite reliability and related reliability evidence reported for latent constructs. Confirmatory factor analysis should evaluate factorial validity, while average variance extracted and latent correlations can contribute to assessment of convergent and discriminant validity. Competing measurement models should also be examined because demonstrating that outside-mind dependency is distinct from AI use, moral disengagement, and authoritarian leadership is essential to establishing the study's theoretical contribution.

After validation of the measurement model, structural equation modelling would test direct and indirect relationships among the principal constructs. Bootstrap confidence intervals would be appropriate for evaluating indirect effects because mediation effects frequently have non-normal sampling distributions, while interaction terms or latent moderated structural equations could test hypothesised moderators. Model interpretation should emphasise effect sizes, confidence intervals, explanatory power, and alternative model comparisons rather than relying exclusively on statistical significance. Cross-sectional structural models would nevertheless remain associational, and temporal language such as “causes” or “leads to” should be avoided until longitudinal or experimental evidence supports causal ordering.

Expected Theoretical Contributions

The first expected contribution is to leadership theory by challenging the assumption that leadership cognition is located solely within the individual leader. The proposed **outside mind** construct reframes leadership as potentially embedded within a sociotechnical cognitive system composed of leaders, advisers, databases, algorithms, generative AI, organisational routines, and institutional knowledge. This does not eliminate individual agency; rather, it creates an empirical question regarding how much substantive cognition and judgement remain with the leader as external systems acquire increasingly important advisory roles. Leadership theory would therefore gain a vocabulary for distinguishing technology *use* from the deeper incorporation of technology into leadership cognition.

A second contribution is the distinction between **cognitive augmentation and cognitive dependency**. Existing discussions sometimes frame AI as either beneficial decision support or a threat to human agency, whereas the proposed model treats these as alternative outcomes of the same underlying process of cognitive extension. This distinction permits researchers to investigate why one leader becomes more reflective and capable when supported by AI while another increasingly defers judgement to algorithmic authority. It consequently moves human-AI leadership research beyond binary questions of adoption towards the quality and structure of cognitive reliance.

A third theoretical contribution concerns the relationship between cognition and moral responsibility. By incorporating Moral Disengagement Theory, the model proposes that technologically mediated leadership may affect not only what decisions leaders make but also how leaders psychologically interpret responsibility for those decisions (Bandura, 1999). A leader who believes that “the algorithm decided” may experience reduced self-sanction despite retaining formal authority over whether the recommendation is implemented. This formulation identifies moral disengagement as a potentially crucial mechanism linking cognitive outsourcing to harmful exercises of organisational power.

The fourth contribution involves connecting AI leadership scholarship explicitly to theories of power and oppression. Much existing research appropriately focuses on efficiency, trust, adoption, fairness, or algorithmic accuracy, but organisational technologies also determine who can act, who can speak, what information is recognised, whose performance is valued, and whether consequential decisions can be challenged. By conceptualising oppression through restrictions on agency, dignity, voice, participation, procedural fairness, and opportunity, the current study provides an ethically serious but empirically testable account of AI-mediated organisational power. It therefore avoids claiming that technology itself oppresses while permitting examination of the institutional conditions through which leaders can use technological systems in oppressive ways.

Expected Practical and Governance Contributions

For organisations, the proposed framework implies that responsible AI governance should examine **cognitive dependency**, not merely cybersecurity, technical accuracy, or regulatory

compliance. A technically accurate system could still contribute to poor leadership if decision-makers cease interrogating recommendations, employees cannot challenge outcomes, or leaders displace responsibility onto technological systems. Governance frameworks should therefore require identifiable human decision ownership, meaningful review of consequential recommendations, documentation of overrides, accessible appeals, employee consultation, and periodic examination of whether AI is reducing managerial critical capacity. Such safeguards would treat human oversight as substantive moral and cognitive engagement rather than the ceremonial presence of a human at the end of an automated process.

Leadership development would likewise need to evolve beyond conventional digital literacy. Leaders working with advanced AI should be trained in **epistemic humility, critical verification, algorithmic limitations, ethical reasoning, responsibility attribution, power awareness, and contestability**, because effective AI leadership requires knowing when *not* to follow a recommendation. Organisations should also evaluate whether productivity incentives unintentionally reward excessive cognitive delegation by privileging speed and quantification over deliberation and employee participation. The central practical implication is therefore that organisations should optimise not simply the quality of artificial intelligence but the quality of the **human-AI cognitive relationship** through which leadership power is exercised.

Expected Discussion

The current study is expected to show that the influence of artificial intelligence on leadership cannot be adequately understood by examining technology adoption alone. The central issue is likely to be the manner in which leaders cognitively incorporate AI and other external resources into judgement, interpretation, and decision-making. Where external systems are used to challenge assumptions, broaden informational awareness, and improve reflection, the outside mind may function as cognitive augmentation and strengthen leadership capability. Conversely, where leaders increasingly defer to external recommendations without critical scrutiny, outside-mind reliance may become cognitive dependency and weaken independent judgement.

A major expected finding is that cognitive dependency may be associated with reduced decision ownership and greater moral disengagement. Bandura's (1999) theory suggests that individuals are more likely to engage in harmful conduct when responsibility is displaced, diffused, or psychologically separated from consequences. In AI-mediated settings, leaders may rationalise difficult or inequitable decisions by attributing authority to “the algorithm,” “the system,” or “the data,” even though they retain responsibility for adopting, interpreting, and implementing technological recommendations. Such a finding would extend moral disengagement theory by demonstrating how technological mediation may create new pathways through which leaders distance themselves from the ethical consequences of organisational decisions.

The study is also expected to demonstrate that oppressive leadership is more likely to emerge when cognitive dependency intersects with concentrated organisational power. AI systems can structure performance evaluation, surveillance, scheduling, recruitment, promotion, and disciplinary processes, thereby embedding leadership authority within technological infrastructures rather than overt interpersonal commands. Lukes' (2005) multidimensional

conception of power is particularly relevant because power may operate by shaping the range of recognised options, legitimate information, and opportunities available to organisational members. Was oppression may become less visibly personal while becoming more systematically embedded in routine organisational processes.

Importantly, the proposed model does not predict that AI use necessarily increase oppressive behaviour. Ethical leadership, critical thinking, AI literacy, transparency, employee participation, and organisational accountability are expected to weaken harmful pathways by ensuring that external recommendations remain subject to human scrutiny and contestation. This distinction is theoretically significant because it avoids technological determinism and recognises that identical AI systems may produce different organisational outcomes depending upon governance arrangements and leadership practice. The quality of the human-AI relationship may therefore be more important than the mere presence or sophistication of artificial intelligence.

The study may also reveal that employee voice is particularly important in constraining oppressive uses of external cognition. When employees can understand, question, appeal, or challenge AI-mediated decisions, technological authority is less likely to become unquestionable organisational power. Conversely, when affected employees lack access to decision criteria or meaningful avenues for contestation, AI-supported recommendations may acquire an aura of inevitability that reduces perceived legitimacy and individual autonomy. These patterns would support contemporary approaches to responsible AI that emphasise transparency, accountability, human oversight, and procedural justice rather than technical accuracy alone.

A further expected contribution concerns the distinction between actual and perceived objectivity. Leaders may place disproportionate confidence in AI-supported recommendations because quantitative systems are often perceived as less emotional, more consistent, or more objective than human judgement. However, computational systems inevitably reflect choices regarding training data, performance objectives, classification thresholds, model design, and organisational priorities. If leaders fail to recognise these embedded human judgements, algorithmic authority may narrow rather than expand critical thinking, thereby increasing the likelihood that biased or incomplete recommendations become institutionalised.

The outside-mind perspective may also help reinterpret established leadership concepts such as distributed leadership and shared cognition. External cognitive influence has always been part of leadership through advisers, committees, professional expertise, and institutional routines, but AI differs in its scale, speed, opacity, and apparent analytical authority. The phenomenon is therefore not simply that cognition has moved outside the individual leader, but that technologically mediated external cognition may increasingly participate in defining what leaders perceive as evidence, risk, performance, and appropriate action. This transformation potentially changes not only decision efficiency but the epistemological basis upon which organisational power is exercised.

Taken together, the expected findings would suggest that the ethical problem of AI leadership is not fundamentally whether machines can possess power or responsibility. Rather, the critical issue is how leaders reorganise their own cognitive and moral relationship to power when

external systems become deeply embedded in decision-making. Responsible AI leadership would therefore require leaders to retain substantive decision ownership, actively interrogate recommendations, recognise uncertainty and bias, and remain answerable for outcomes affecting others. The outside-mind framework consequently offers a way to connect leadership cognition, artificial intelligence, ethics, and organisational power within a single theoretically coherent model.

Future Research

Future research should first establish whether the proposed outside-mind construct is psychometrically distinct from related concepts such as automation bias, technology acceptance, delegation, distributed cognition, and decision dependence. Longitudinal studies would then be required to determine whether sustained reliance on AI changes leadership judgement over time or whether leaders who are already more dependent on external guidance are simply more likely to adopt AI-intensive decision practices. Experimental designs could also manipulate the apparent authority, transparency, or certainty of AI recommendations to test whether these features alter leaders' willingness to override systems and accept personal responsibility. Such studies would provide stronger causal evidence than cross-sectional associations alone.

Further research should examine organisational context because the effects of the outside mind are unlikely to be uniform across industries, cultures, or leadership levels. High-risk sectors such as healthcare, policing, finance, education, and human-resource management may produce different ethical dynamics from settings where AI primarily supports routine administrative decisions. Cross-cultural studies should also examine power distance, institutional trust, regulatory traditions, and norms of employee participation because these factors may substantially influence whether AI-mediated authority is questioned or accepted. Comparative research across countries would therefore help determine whether the proposed framework generalises beyond specific organisational environments.

Qualitative research also be important for understanding how leaders actually experience external cognitive influence. Interviews, ethnographies, and decision diaries could explore when leaders feel empowered by AI, when they feel dependent upon it, and how they assign responsibility when technologically assisted decisions produce harmful consequences. Employee perspectives should be investigated simultaneously because the experience of technological oppression may differ substantially from leaders' perceptions of their own fairness or accountability. Combining these perspectives would strengthen understanding of the subjective and relational dimensions of AI-mediated power.

Future research should also examine interventions designed to preserve human judgement and moral responsibility. Experimental or field studies could test whether AI-literacy training, structured critical-reflection protocols, mandatory explanation requirements, employee appeals, algorithmic audits, or human-override procedures reduce cognitive dependency and moral disengagement. Research could additionally evaluate whether organisations that reward thoughtful disagreement with AI produce stronger ethical outcomes than organisations that primarily reward speed, compliance, and predictive efficiency. Such work would move the

literature from identifying risks towards establishing evidence-based safeguards for responsible AI-supported leadership.

Strengths and Limitations

The current study possesses several conceptual strengths. It integrates traditionally separate literatures on extended cognition, distributed cognition, leadership, moral disengagement, artificial intelligence, and organisational power, thereby addressing a fragmented area of scholarship. The distinction between cognitive augmentation and cognitive dependency also provides a more nuanced account than approaches that classify AI simply as beneficial or harmful. Finally, the multisource design and proposed employee-level assessment of oppression strengthen construct validity by avoiding exclusive reliance on leaders' self-perceptions.

Several limitations nevertheless require acknowledgement. The outside mind is a newly proposed construct and therefore lacks an established measurement tradition, meaning that extensive scale validation was necessary before substantive claims can be considered definitive. Cross-sectional data would also limit causal inference, while self-reported measures of cognitive reliance and moral responsibility may be affected by social desirability or limited introspective accuracy. Moreover, the meaning and severity of oppression are context-dependent, requiring careful operationalisation to avoid conflating legitimate managerial authority with systematic restrictions on agency, dignity, fairness, and participation.

Implications for Leadership Theory

The proposed Outside Mind-Leadership-Power Model has important implications for the conceptual foundations of leadership scholarship because it challenges the assumption that leadership behaviour originates principally from cognition located within an identifiable individual. Traditional approaches commonly attribute vision, judgement, ethical reasoning, strategic choice, and responsibility to the leader, even where contextual and follower influences are acknowledged. However, Extended Mind Theory suggests that cognitive processes can incorporate external resources when those resources become functionally integrated into thinking and action (Clark & Chalmers, 1998). The increasing reliance on generative AI therefore requires leadership scholarship to consider whether the effective cognitive unit responsible for some leadership decisions is becoming a **human-AI cognitive system rather than the unaided individual leader**.

This reconceptualisation does not diminish the importance of the leader but changes the theoretical meaning of leadership agency. A leader may retain formal authority while increasingly depending upon external systems to identify problems, generate alternatives, forecast consequences, evaluate employees, recommend actions, and formulate explanations for decisions. Distributed Cognition Theory is useful here because cognition can be understood as occurring across people, artefacts, and environments rather than being confined to individual minds (Hutchins, 1995). The outside-mind perspective extends this reasoning by examining when distributed cognition ceases to function merely as collaboration and begins to redistribute the **effective location of judgement and cognitive authority**.

An important theoretical distinction consequently arises between **formal decision authority** and **effective cognitive authority**. Formal authority refers to the recognised organisational power to make a decision, whereas effective cognitive authority concerns the source that substantially determines how the problem is understood and which decision appears appropriate. A manager may formally approve a promotion, dismissal, workload allocation, or disciplinary decision while an algorithm has effectively determined the ranking, classification, prediction, or recommendation upon which that decision depends. Leadership research that examines only the final human decision-maker may therefore underestimate the extent to which leadership cognition has already migrated into external technological infrastructures.

This argument suggests that contemporary leadership theory should distinguish among **decision authorship, decision approval, and decision responsibility**. A leader may approve an externally generated recommendation without being its substantive cognitive author, yet organisational and ethical responsibility may remain with the leader who possesses the authority to implement or reject it. The distinction becomes particularly important when AI-generated recommendations have adverse consequences because leaders may perceive reduced responsibility for decisions they did not cognitively originate. The outside-mind framework therefore expands leadership theory by showing that technological mediation can separate cognitive authorship from formal authority without necessarily eliminating moral accountability.

The Outside Mind and the Transformation of Organisational Power

The outside mind also provides a different perspective on organisational power because AI allows influence to be exercised indirectly through informational architectures. French and Raven's (1959) classical taxonomy demonstrates that organisational influence can arise through legitimate, reward, coercive, expert, and referent power, while subsequent scholarship has shown that power also operates structurally by shaping available choices and recognised interests (Lukes, 2005). AI potentially expands expert power because algorithmic systems can acquire authority from their perceived analytical sophistication, predictive capacity, and apparent objectivity. When leaders and employees assume that computational recommendations possess superior epistemic legitimacy, technological expertise may become an indirect source of organisational authority.

This creates what may be termed **delegated algorithmic power**, whereby leaders retain institutional authority but permit technological systems to exercise substantial influence over the criteria through which employees are evaluated or organisational resources distributed. Such power can remain difficult to recognise because the algorithm does not necessarily issue commands directly; rather, it establishes classifications, scores, rankings, predictions, or recommendations that subsequently shape human decisions. The resulting organisational hierarchy may therefore include technological systems as influential intermediaries between those possessing formal authority and those subjected to its consequences. This represents an important extension of traditional power analysis because the mechanism exercising influence may possess neither formal organisational status nor moral agency.

AI may further generate **epistemic asymmetry** between leaders and followers. Leaders may possess access to predictive analytics, employee classifications, performance dashboards, and AI-generated interpretations unavailable to employees, while those employees may have limited knowledge of how these systems evaluate their behaviour. This informational inequality becomes a power issue when individuals cannot understand the basis of decisions affecting their employment, remuneration, promotion, scheduling, workload, or professional opportunities. Transparency is not merely a technical characteristic of responsible AI but an important determinant of the distribution of organisational power.

The concept of **contestability** becomes particularly important in this context. An employee subjected to an adverse human decision may at least theoretically question the reasoning of the decision-maker, but contesting an algorithmically mediated decision becomes more difficult when neither employee nor manager can adequately explain how the output was produced. Organisational accountability therefore requires more than keeping a nominal human “in the loop”; the responsible human must possess sufficient authority, knowledge, and insight to interrogate and override the system. Where meaningful contestability disappears, technological recommendations risk becoming functionally authoritative regardless of whether formal decision-making technically remains human.

From Outside Mind to Oppressive Power

The proposed theory suggests that oppression becomes possible when external cognitive authority combines with asymmetrical organisational power and diminished contestability. Oppression should nevertheless remain conceptually distinct from ordinary hierarchy because organisations inevitably require coordination, authority, performance standards, and legitimate accountability. The defining issue is whether power becomes systematically organised in ways that diminish the agency, dignity, voice, autonomy, fair treatment, or meaningful participation of individuals subjected to it. This definition permits empirical investigation without presuming that every unequal organisational relationship is inherently oppressive.

AI-mediated oppression may be especially difficult to recognise because it can operate through apparently neutral procedures rather than explicit hostility. An algorithm could systematically determine who receives undesirable shifts, heightened surveillance, lower performance ratings, fewer advancement opportunities, or greater disciplinary scrutiny without any manager explicitly expressing discriminatory or oppressive intentions. The absence of visible interpersonal hostility does not necessarily eliminate structural disadvantage if the system repeatedly distributes burdens and opportunities unequally. This observation makes the outside-mind framework particularly relevant to contemporary organisations because technological mediation can potentially separate oppressive consequences from identifiable oppressive intentions.

Moral disengagement provides a plausible psychological mechanism connecting these structural processes to individual leadership behaviour. Bandura (1999) demonstrated that displacement and diffusion of responsibility can weaken self-regulatory mechanisms that would otherwise constrain harmful conduct. A leader who considers an adverse decision to be “the algorithm’s recommendation” may experience less personal moral conflict than when making precisely the

same decision independently. The important theoretical proposition is therefore that AI may sometimes change not only **what leaders decide but how morally close leaders feel to the consequences of what they decide**.

This process can be conceptualised as **algorithmic moral distancing**: the psychological separation of a human decision-maker from the moral consequences of a decision because substantial cognitive authorship is attributed to an external technological system. Algorithmic moral distancing would represent a specific technologically mediated pathway within the broader construct of moral disengagement rather than an entirely separate moral theory. Its presence could be measured through leaders' attribution of responsibility, willingness to override recommendations, perceived personal accountability, and emotional responses to adverse outcomes. Establishing this construct empirically would constitute an important extension of the present study.

A Human-AI Leadership Continuum

Rather than treating AI involvement as dichotomous, the proposed framework suggests that contemporary leadership may exist along a continuum of cognitive externalisation. At one end is **human-dominant cognition**, where leaders obtain external information but independently interpret evidence and formulate decisions; at the next level is **AI-assisted cognition**, where technological systems provide recommendations that remain critically evaluated by human decision-makers. A third level involves **AI-dependent cognition**, where external recommendations become the dominant basis for judgement, and leaders rarely depart from them. At the extreme lies **nominal human leadership**, where the leader formally approves decisions that are substantively formulated, structured, and determined by technological systems.

This continuum provides a more precise basis for empirical research than simply comparing organisations that “use AI” with those that do not. Two leaders may use the same generative AI platform daily while occupying completely different positions on the continuum because one actively challenges its recommendations whereas the other routinely accepts them. Frequency of use is therefore theoretically distinct from depth of cognitive externalisation. Future measurement should consequently assess **how AI is incorporated into cognition**, rather than merely how frequently leaders access technological systems.

A similar continuum can be proposed for responsibility. At one extreme, leaders maintain **full moral ownership**, explicitly accepting responsibility for technologically assisted decisions and their foreseeable consequences; intermediate positions involve shared or ambiguous responsibility among leaders, organisations, developers, and technological systems. At the opposite extreme lies **responsibility displacement**, in which human actors attribute adverse outcomes primarily to technology. The most ethically problematic configuration would therefore combine high cognitive dependency, low decision ownership, high moral disengagement, weak accountability, and substantial organisational power.

These propositions strengthen the explanatory potential of the conceptual framework because they replace a simplistic AI-versus-human distinction with multidimensional relationships. The fundamental variables become cognitive externalisation, critical scrutiny, decision ownership,

moral responsibility, power asymmetry, transparency, and contestability. Leadership effectiveness and ethicality can then be examined as outcomes of particular configurations among these dimensions. Such an approach may prove especially useful as generative and agentic AI systems acquire greater capacity to perform tasks previously understood as inherently managerial.

Broader Social Implications

Although the current study is initially situated within organisational leadership, the outside-mind framework may eventually have applications beyond conventional workplaces. Political leaders, educational administrators, healthcare executives, military commanders, public-sector officials, and other authority figures increasingly operate within technologically mediated information environments. In each setting, AI can influence what information reaches decision-makers, how problems are classified, which risks receive priority, and which interventions are presented as optimal. The broader societal question is therefore whether technologically extended cognition changes how authority itself is exercised.

This question becomes especially significant when decisions affect populations that possess comparatively little power to challenge institutional authority. If AI systems are incorporated into welfare administration, employment selection, policing, healthcare allocation, education, or other consequential domains, individuals may experience decisions generated through complex cognitive networks rather than identifiable human deliberation. Lukes' (2005) analysis of power suggests that domination can operate by structuring the circumstances within which decisions and alternatives become possible. The outside mind therefore provides a potential conceptual bridge connecting leadership scholarship with wider debates concerning technological governance and institutional power.

The framework nevertheless rejects the proposition that retaining exclusively human decision-making necessarily guarantees justice. Human leadership has historically produced prejudice, discrimination, coercion, exploitation, and oppression without artificial intelligence, while appropriately designed technological systems may sometimes expose inconsistencies or reduce particular forms of arbitrary judgement. The normative objective should consequently not be to preserve human cognition merely because it is human. Instead, the objective should be to construct **human-AI cognitive arrangements that improve knowledge while preserving accountability, dignity, contestability, and moral agency.**

The central principle emerging from the study can therefore be expressed succinctly: **the extension of the leader's mind should not produce the contraction of the follower's agency.** External intelligence becomes ethically defensible when it expands leaders' capacity to understand complex realities while preserving the ability of affected individuals to be heard, respected, informed, and treated fairly. It becomes problematic when increased cognitive capacity at the top of an organisation simultaneously produces intensified surveillance, diminished autonomy, opaque authority, or reduced opportunities for those below to challenge decisions. This proposition could ultimately become the normative foundation of the Outside Mind-Leadership-Power Model.

Proposed Theoretical Propositions

For subsequent conceptual publication, the framework can be condensed into a series of theoretical propositions. First, increasing external cognitive reliance does not necessarily diminish leadership autonomy; its effects depend upon whether reliance takes the form of augmentation or dependency. Second, cognitive dependency increases the likelihood of responsibility displacement when leaders perceive technological recommendations as possessing superior epistemic authority. Third, responsibility displacement and moral disengagement increase the likelihood that organisational power is exercised without adequate consideration of consequences for employee agency, dignity, and voice.

Fourth, the relationship between outside-mind dependency and oppressive leadership was stronger under conditions of high power asymmetry, low transparency, weak accountability, and limited employee contestability. Fifth, ethical leadership, critical thinking, AI literacy, meaningful human oversight, and employee voice reduce the probability that cognitive externalisation produces oppressive outcomes. Sixth, high levels of AI use combined with strong critical scrutiny and retained decision ownership may improve leadership without increasing oppression. Finally, the ethical quality of AI-mediated leadership depends less upon whether cognition extends beyond the individual mind than upon **how cognitive authority, organisational power, and moral responsibility are distributed within the resulting human-AI system.**

These propositions elevate the manuscript beyond a study of AI adoption and position the **Outside Mind** as a potentially original theoretical contribution to leadership scholarship. The next developmental stage should be a dedicated section on “**Development of the Outside Mind Theory of Leadership**”, in which the construct is formally defined, its assumptions and boundaries are established, and it is differentiated from Extended Mind Theory, Distributed Cognition, algorithmic management, shared leadership, and AI-assisted decision-making. That would make the manuscript substantially stronger as a theory-building paper rather than simply an empirical proposal.

Development of the Outside Mind Theory of Leadership

The **Outside Mind Theory of Leadership** proposes that leadership cognition may extend beyond the individual leader into a wider network of human, technological, informational, and institutional resources that participate meaningfully in judgement and decision-making. The theory builds on Extended Mind Theory and Distributed Cognition but moves beyond those traditions by focusing specifically on leadership authority, moral agency, and the consequences of externally supported cognition for others (Clark & Chalmers, 1998; Hutchins, 1995). Its central assumption is that external cognitive resources can become functionally integrated into leadership processes when they are repeatedly used to frame problems, interpret information, generate alternatives, evaluate consequences, and justify decisions. The theory therefore treats the leader not as an isolated cognitive actor but as part of a potentially distributed decision system whose ethical quality depends upon how cognitive authority and responsibility are organised.

A second core assumption is that external cognition can produce qualitatively different outcomes depending upon whether it functions as **augmentation** or **dependency**. Cognitive augmentation occurs when external resources increase informational breadth, expose leaders to alternative interpretations, and strengthen reflective judgement without displacing personal responsibility. Cognitive dependency occurs when leaders increasingly defer to external systems, advisers, or AI-generated recommendations in ways that reduce independent scrutiny and weaken decision ownership. This distinction is fundamental because it prevents the theory from assuming that greater use of external cognition necessarily produces harmful leadership.

The theory also assumes that leadership power and cognition cannot be analysed independently. Decisions regarding recruitment, promotion, surveillance, performance assessment, discipline, workload, compensation, and resource allocation affect the autonomy and opportunities of organisational members, meaning that cognition is translated into consequential exercises of power. When external systems substantially influence these decisions, the structure of power becomes partly embedded within the informational and technological architecture through which decisions are produced. Accordingly, Outside Mind Theory examines not only who formally authorises a decision but also which cognitive system effectively determines what options appear valid or legitimate.

A fourth assumption concerns **moral responsibility**. The theory proposes that technological or external mediation does not eliminate human responsibility merely because cognitive authorship is distributed across several actors or systems. Leaders who possess the authority to accept, reject, or implement externally generated recommendations remain morally implicated in the consequences of those choices, particularly when decisions materially affect others. The ethical risk emerges when distributed cognition is accompanied by responsibility displacement, allowing leaders to treat harmful outcomes as products of “the system” rather than as consequences of human governance and decision-making.

Core Constructs of Outside Mind Theory

The first core construct is **external cognitive incorporation**, defined as the degree to which resources outside the biological individual become functionally embedded within leadership thinking. These resources may include advisers, consultants, committees, organisational routines, databases, predictive analytics, decision-support systems, and generative AI. External cognitive incorporation differs from ordinary information gathering because the external resource becomes an active component in interpreting situations and determining what action should follow. The greater the cognitive role of the external resource, the more leadership judgement becomes distributed across a sociotechnical network.

The second construct is **cognitive authority**, referring to the degree of epistemic influence attributed to an external source. Leaders may treat recommendations differently depending upon whether they perceive the source as advisory, expert, authoritative, objective, or effectively determinative. AI systems may acquire unusually high cognitive authority because quantitative outputs can appear precise, impartial, and analytically superior even where uncertainty, bias, or model limitations remain substantial. The theory therefore predicts that perceived cognitive

authority, rather than frequency of technology use alone, is a critical determinant of cognitive dependency.

The third construct is **decision ownership**, which captures the extent to which leaders regard themselves as substantively responsible for the decisions they implement. High decision ownership involves understanding the basis of a recommendation, critically evaluating alternatives, accepting responsibility for consequences, and being willing to justify the final course of action. Low decision ownership occurs when leaders merely approve or transmit externally generated decisions without meaningful examination of their assumptions or likely effects. Decision ownership therefore represents an important bridge between cognitive externalisation and moral agency.

The fourth construct is **algorithmic moral distancing**, defined as the psychological separation of leaders from the ethical consequences of decisions because those decisions are substantially attributed to technological systems. This mechanism is conceptually grounded in Moral Disengagement Theory, particularly displacement and diffusion of responsibility (Bandura, 1999). A leader may experience less moral discomfort when an adverse employment decision is framed as algorithmically recommended than when the identical decision is understood as personally authored. Outside Mind Theory predicts that moral distancing was strongest when cognitive dependency is high and organisational accountability is weak.

Boundary Conditions of the Theory

Outside Mind Theory does not claim that all external influence constitutes an outside mind in the theoretically meaningful sense. Occasional consultation, passive information retrieval, or incidental technological assistance would not qualify unless the external resource becomes recurrently incorporated into leadership cognition and materially influences judgement. The theory therefore requires a threshold of functional integration rather than simple exposure to external information. This boundary condition protects the construct from becoming so broad that every form of advice or technology use is classified as outside-mind leadership.

The theory also does not presume that human cognition is inherently more ethical or accurate than AI-supported cognition. Human leaders can display prejudice, poor judgement, self-interest, and oppressive behaviour independently of technological assistance, while appropriately governed AI may reduce inconsistency or reveal information that improves fairness. The normative concern is therefore not the preservation of unaided human judgement as an end in itself but the preservation of critical scrutiny, responsibility, dignity, and contestability within distributed decision systems. Outside Mind Theory is thus compatible with substantial AI use provided meaningful human moral agency remains intact.

A further boundary condition concerns the distinction between **oppression and legitimate organisational control**. Organisations require rules, hierarchy, accountability, and coordination, and not every restriction on discretion constitutes oppressive leadership. Oppression arises when power is exercised systematically in ways that substantially diminish voice, autonomy, dignity, fair treatment, participation, or access to opportunities without adequate justification or

meaningful avenues for challenge. This distinction ensures that the theory evaluates the quality and consequences of power rather than treating authority itself as inherently problematic.

The theory is also most applicable to decisions that carry meaningful consequences for other people. Routine administrative uses of AI with minimal human impact may involve distributed cognition but are less relevant to the leadership-power mechanisms proposed here. The framework becomes most analytically important where external cognition influences consequential decisions involving employment, surveillance, discipline, promotion, compensation, allocation of opportunities, or access to organisational resources. The magnitude of potential harm therefore forms an important contextual boundary for empirical application.

Distinguishing Outside Mind Theory from Related Perspectives

Outside Mind Theory differs from **Extended Mind Theory** primarily in its object of explanation. Extended Mind Theory addresses the philosophical question of whether cognitive processes may extend beyond the biological individual into external artefacts or environments (Clark & Chalmers, 1998). Outside Mind Theory accepts this possibility but focuses specifically on leadership, decision authority, moral responsibility, and the social consequences of distributed cognition. It therefore transforms a theory of cognition into a leadership framework concerned with power and ethical accountability.

The framework also differs from **Distributed Cognition**, although it draws substantially from that tradition. Distributed Cognition examines how cognitive processes occur across people, artefacts, and environments, particularly in complex task systems (Hutchins, 1995). Outside Mind Theory adds a normative and hierarchical dimension by asking what happens when distributed cognition is embedded within relationships where one actor possesses authority over another. Its distinctive contribution lies in connecting distributed cognitive processes to asymmetrical power, employee agency, moral disengagement, and oppression.

Outside Mind Theory is likewise distinct from **algorithmic management**. Algorithmic management describes organisational practices in which software systems automate or support activities traditionally undertaken by human managers, including monitoring, allocation, evaluation, and coordination. Outside Mind Theory examines a deeper cognitive question: how does repeated dependence on such systems reshape what leaders think, how they reason, and how responsibility is experienced? Algorithmic management can therefore be treated as one organisational context in which outside-mind processes may emerge rather than as an equivalent concept.

The theory also differs from **shared leadership** and related collective leadership perspectives. Shared leadership concerns the distribution of leadership influence among multiple human actors within groups or teams, often emphasising collaborative influence rather than hierarchical authority. Outside Mind Theory includes human advisers and networks but also incorporates non-human cognitive resources and focuses on the possibility that cognitive authority becomes distributed without equivalent distribution of formal power or responsibility. This asymmetry between distributed cognition and concentrated authority is one of the theory's most distinctive propositions.

Fundamental Propositions of Outside Mind Theory

The first proposition is that the degree of external cognitive incorporation increase as leaders confront greater informational complexity, uncertainty, time pressure, and access to advanced AI systems. However, increased incorporation not necessarily reduce leadership quality because external cognition can operate through either augmentation or dependency. Leaders with strong critical-thinking skills and high AI literacy are expected to use external systems as resources for comparison and reflection rather than as substitutes for judgement. The effect of outside-mind reliance is therefore conditional rather than inherently positive or negative.

The second proposition is that **cognitive dependency increases when perceived cognitive authority exceeds critical scrutiny**. Leaders are more likely to defer to external recommendations when technological outputs are perceived as objective, precise, or institutionally endorsed and when opportunities for independent verification are limited. Repeated deference may gradually normalise the external system as the default source of valid judgement. Over time, this process may shift leadership behaviour from AI-assisted decision-making towards nominal human approval of externally determined decisions.

The third proposition is that cognitive dependency reduce decision ownership unless accountability structures explicitly require leaders to retain responsibility. When leaders view themselves as executors rather than authors of externally generated decisions, the psychological distance between action and consequence may increase. Moral disengagement mechanisms such as displacement of responsibility can then become more accessible, particularly where organisational culture encourages reliance on technological recommendations. Thus, the pathway from outside mind to harmful leadership is expected to operate partly through diminished decision ownership.

The fourth proposition is that **oppressive outcomes become more likely when low decision ownership coincides with high organisational power and weak employee contestability**. Leaders who can implement consequential recommendations without explaining their basis or allowing meaningful challenge possess a structurally advantageous position over those subjected to the decisions. If algorithmic or external recommendations are treated as unquestionable, employees may lose practical opportunities to influence decisions that affect their work and careers. Outside Mind Theory therefore predicts that the distribution of contestability is as important as the distribution of cognitive authority.

Protective Mechanisms within Outside Mind Theory

Ethical leadership is expected to function as one of the strongest protective mechanisms. Ethical leaders are more likely to consider consequences for affected individuals, communicate decision rationales, accept responsibility, and challenge recommendations that conflict with fairness or human dignity. These behaviours can preserve moral agency even when extensive external cognitive support is present. Ethical leadership therefore interrupts the pathway between cognitive dependency and oppressive power.

AI literacy constitutes a second protective mechanism because leaders who understand model limitations, uncertainty, data bias, and the difference between prediction and judgement are less likely to attribute excessive authority to technological outputs. AI literacy should encourage leaders to ask what data informed a recommendation, what assumptions underpin it, what information may be missing, and which populations may be disadvantaged by its use. Such reflective practices convert technological assistance from an authority relationship into an evidence relationship. Accordingly, AI literacy is expected to increase augmentation while reducing dependency.

Organisational accountability represents another critical safeguard. Clear responsibility allocation, audit trails, mandatory human review, documented rationales, appeal mechanisms, and periodic algorithmic audits can prevent leaders from displacing responsibility onto technological systems. These governance structures also communicate institutionally that AI recommendations remain subject to human judgement rather than constituting autonomous organisational authority. Accountability therefore protects both employees and leaders by clarifying where responsibility resides.

Employee voice and contestability provide an additional structural safeguard. When employees can question decisions, access explanations, identify errors, and appeal harmful outcomes, leaders receive information that may expose weaknesses in externally generated recommendations. Voice also counterbalances informational asymmetry by preventing decision-making systems from becoming entirely one-directional. Outside Mind Theory consequently treats employee participation not merely as a positive organisational practice but as a mechanism that constrains potentially oppressive uses of technologically mediated power.

Formal Statement of the Theory

Outside Mind Theory of Leadership can be stated formally as follows: **leadership behaviour and the exercise of organisational power are shaped by the degree and manner in which external cognitive resources become incorporated into leadership judgement.** External cognition produces beneficial outcomes when it augments critical reflection while preserving decision ownership, moral agency, transparency, accountability, and employee contestability. It produces increased risk of oppressive outcomes when cognitive dependency combines with responsibility displacement, moral disengagement, concentrated power, opaque decision processes, and weak opportunities for those affected to challenge decisions. The theory therefore predicts that the ethical consequences of AI-mediated leadership emerge from the architecture of the human-external cognition relationship rather than from technological use alone.

The theory's central normative proposition is equally important: **leadership cognition may extend beyond the individual mind, but moral responsibility should not extend beyond identifiable human accountability.** Leaders may legitimately distribute cognitive work across humans and technologies while still retaining responsibility for how that work is interpreted and translated into action. Organisations should therefore distinguish distributed intelligence from distributed accountability because the latter can create conditions in which responsibility

becomes so fragmented that no actor meaningfully owns harmful outcomes. This distinction provides the ethical core of the proposed theory.

A second normative proposition is that increases in leadership cognitive capacity should not be purchased through reductions in follower agency. AI may enable leaders to know more, predict more, monitor more, and decide more quickly, but technological capability does not itself establish ethical legitimacy for expanded control. Responsible leadership requires that enhanced cognitive power be balanced by transparency, proportionality, employee voice, procedural fairness, and meaningful possibilities for challenge. This principle connects technological governance directly to established concerns regarding dignity and legitimate organisational authority.

Finally, Outside Mind Theory positions AI-mediated leadership as an emerging **sociotechnical form of leadership** rather than simply traditional leadership supplemented by a new tool. As external systems participate more deeply in cognition, distinctions among adviser, information source, decision-support system, and quasi-decision-maker become increasingly important. Leadership scholarship must therefore examine not only leader traits and behaviours but also the cognitive infrastructures through which those traits and behaviours are translated into decisions. This theoretical shift offers a foundation for a new research programme on cognition, technology, morality, and organisational power in the age of artificial intelligence.

Conclusion

The increasing incorporation of artificial intelligence into organisational decision-making requires leadership scholarship to reconsider where cognition, judgement, and responsibility are located. This study proposes the concept of the **outside mind** to describe the network of external human, technological, informational, and institutional resources that become integrated into leadership thought and decision-making. The framework distinguishes cognitive augmentation, in which external intelligence strengthens human judgement, from cognitive dependency, in which leaders increasingly surrender critical evaluation and decision ownership to external systems. This distinction provides a foundation for understanding why the same technologies may support responsible leadership in one context while facilitating oppressive organisational control in another.

The central theoretical argument is that oppression does not arise simply because AI participates in leadership decisions. Harm becomes more likely when technological authority combines with weak critical scrutiny, moral disengagement, limited transparency, concentrated organisational power, and reduced employee voice. Under these conditions, AI-mediated decisions may constrain autonomy, suppress participation, institutionalise inequity, or obscure responsibility while appearing neutral or technically inevitable. Conversely, ethical leadership, AI literacy, accountability, transparency, and meaningful contestability can preserve the benefits of external cognition while maintaining human moral agency.

The study therefore shifts the central debate from whether leaders should use AI to **how leaders should think with AI while remaining responsible for the power they exercise over others**. Leadership beyond the individual mind need not imply leadership beyond human responsibility;

rather, increasingly distributed cognition makes the preservation of responsibility even more important. Organisations should therefore design AI governance around not only technical performance but also decision ownership, ethical reflection, employee participation, and the capacity to challenge technologically mediated outcomes. In this way, the outside-mind framework offers a theoretically grounded foundation for future research on responsible leadership in increasingly AI-mediated organisations.

References

- Atianashie, M. A., Kuffour, M. K., Kyiewu, B., & Serwaa, P. (2025). Responsible AI for AI sustainable future: Governance, ethics, and the reality behind the promise. *Journal of Information Technology, Cybersecurity, and Artificial Intelligence*, 2(2), 1-19. <https://doi.org/10.70715/jitcai.2025.v2.i2.012>
- Baciu, C., Mohan, G., & Wilson, J. (2026). Algorithmic scholarship and academic evaluation: Governance misalignment in the age of generative AI. *Frontiers in Artificial Intelligence*, 9, 1862980. <https://doi.org/10.3389/frai.2025.1862980>
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193-209.
- Bass, B. M., & Bass, R. (2008). *The Bass handbook of leadership: Theory, research, and managerial applications* (4th ed.). Free Press.
- Burns, J. M. (1978). *Leadership*. Harper & Row.
- Chigbu, B. I. (2026). Algorithmic management in the global gig economy: An interdisciplinary systematic literature review and critical discourse analysis. *Frontiers in Sociology*, 11, 1743445. <https://doi.org/10.3389/fsoc.2025.1743445>
- Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7-19. <https://doi.org/10.1093/analys/58.1.7>
- Collins, P. (2023). Worker voice and algorithmic management in post-Brexit Britain. *Transfer: European Review of Labour and Research*.
- Doellgast, V., et al. (2023). Negotiating limits on algorithmic management in digitalised services. *Transfer: European Review of Labour and Research*.
- Favela, L. H., Amon, M. J., Lobo, L., & Chemero, A. (2021). Empirical Evidence for Extended Cognitive Systems. *Cognitive Science*, 45(11), e13060. <https://doi.org/10.1111/cogs.13060>
- French, J. R. P., Jr., & Raven, B. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in social power* (pp. 150-167). Institute for Social Research.
- Greif, H. (2015). What is the extension of the extended mind? *Synthese*, 194(11), 4311. <https://doi.org/10.1007/s11229-015-0799-9>
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I.

- (2018). *Mixed Methods Appraisal Tool (MMAT), version 2018: User guide*. McGill University.
- Hossain, S., Fernando, M., & Akter, S. (2025). The influence of artificial intelligence-driven capabilities on responsible leadership: A future research agenda. *Journal of Management & Organization*, 31(5).
- Hutchins, E. (1995). *Cognition in the wild*. MIT Press.
- International Labour Organization. (2025). *Algorithmic management in the workplace*. International Labour Organization.
- Jafari, A. A., Ozcinar, C., & Anbarjafari, G. (2026). *Responsible AI: The good, the bad, the AI* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2601.21095>
- Kuchař, P. (2025). Competition as socially extended cognition: A Hayekian perspective on market knowledge and institutional evolution. *The Review of Austrian Economics*. <https://doi.org/10.1007/s11138-025-00695-1>
- Lippert, I., Alizadeh, A., Tarafdar, M., Mohlmann, M., Benlian, A., Parent-Rochelleau, X., & Stein, M.-K. (2026). One decade of algorithmic management research. *Business & Information Systems Engineering*, 68, 959-969. <https://doi.org/10.1007/s12599-026-00995-1>
- Loock, L. (2025). The extracted mind. *Synthese* 205, 133. <https://doi.org/10.1007/s11229-025-04962-3>
- Lukes, S. (2005). *Power: A radical view* (2nd ed.). Palgrave Macmillan.
- Milanez, A. (2025). *Algorithmic management in the workplace*. OECD Publishing.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71.
- Plantinga, P. (2024). Responsible artificial intelligence in Africa: Towards policy learning. *Data & Policy*.
- Sprevak, M. (2009). Extended cognition and functionalism. *The Journal of Philosophy*, 106(9), 503-527.
- Sullivan, R., Veen, A., & Riemer, K. (2024). Furthering engaged algorithmic management research: Surfacing foundational positions through a hermeneutic literature analysis. *Information and Organization*, 34(4), 100528. <https://doi.org/10.1016/j.infoandorg.2024.100528>
- Telakivi, P. (2020). *Extending the extended mind: From cognition to consciousness* [Doctoral dissertation, University of Helsinki]. University of Helsinki. <https://helda.helsinki.fi/server/api/core/bitstreams/57445211-bcf5-4d66-bc67-345e5b2ea52e/content>
- Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, 45.

- Tigard, D. W. (2021). Artificial moral responsibility: How we can and cannot hold machines responsible. *Cambridge Quarterly of Healthcare Ethics*, 30, 435-447.
- Wilson, C. E., de Vries, J. L., & Nguyen, E. T. (2025). Extended cognition and its influence on humanistic inquiry: A framework for interdisciplinary integration. *Future-Humanities and Technological Integration*, 1(2). <https://fupress.org/journal/HTI/index.php/journal/article/view/76>
- Yukl, G. (2013). *Leadership in organizations* (8th ed.). Pearson.

Appendix A

Proposed Outside Mind Leadership Questionnaire (OMLQ)

Purpose. The proposed Outside Mind Leadership Questionnaire is designed to operationalise the principal dimensions emerging from the Outside Mind Theory of Leadership for subsequent quantitative validation. The instrument measures external cognitive incorporation, cognitive augmentation, cognitive dependency, leadership decision-making autonomy, decision ownership, moral agency and disengagement, organisational power, employee agency, protective governance conditions, and empowering versus oppressive leadership. Unless otherwise indicated, responses would be recorded on a **5-point Likert scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, and 5 = Strongly agree**. Higher subscale scores indicate greater levels of the construct being measured, while reverse-coded items are identified by **(R)**.

Table A1: Proposed Dimensions, Operational Definitions, and Measurement of the Outside Mind Leadership Model

Dimension	Operational definition	Proposed items	Measurement/ interpretation
External Cognitive Influence/Outside Mind (ECI)	Extent to which external human, technological, informational, and institutional resources become incorporated into leadership cognition	5	Higher = greater external cognitive incorporation
Cognitive Augmentation (CA)	Extent to which external resources improve judgement while preserving critical human evaluation	5	Higher = stronger cognitive augmentation
Cognitive Dependency (CD)	Extent to which leadership judgement becomes dependent upon external or AI-generated recommendations	5	Higher = greater cognitive dependency
Decision-Making Autonomy/Critical Scrutiny (DMA)	Capacity to independently evaluate, question, modify, or reject external recommendations	5	Higher = greater retained cognitive autonomy
Decision Ownership (DO)	Extent to which leaders accept authorship and responsibility for AI-assisted decisions	4	Higher = greater decision ownership
Moral Agency (MA)	Extent to which ethical responsibility remains central to externally assisted decisions	4	Higher = stronger moral agency
Moral Disengagement/Algorithmic Moral Distancing (MD)	Psychological displacement or diffusion of responsibility to AI, algorithms, data, or organisational systems	5	Higher = greater moral disengagement

Exercise of Organisational Power (OP)	Extent to which external cognitive systems contribute to control, surveillance, evaluation, and consequential decisions affecting employees	5	Higher = greater AI-mediated organisational control
Employee Voice and Contestability (EVC)	Employees' ability to question, explain, appeal, or influence AI-mediated decisions	5	Higher = greater voice and contestability
Employee Autonomy, Dignity, Fairness and Participation (EADFP)	Extent to which technologically mediated leadership preserves employee agency, respectful treatment, fairness, and participation	6	Higher = stronger employee agency and procedural protection
AI Governance and Accountability (AGA)	Presence of transparency, human oversight, accountability, explainability, and governance safeguards	5	Higher = stronger governance protection
Empowering/Responsible Leadership (ERL)	Use of external cognition to strengthen leadership while maintaining accountability and employee agency	5	Higher = more responsible/empowering leadership
Oppressive Leadership (OL)	Systematic use of leadership power in ways that diminish employee agency, voice, dignity, fairness, autonomy, or meaningful contestability	6	Higher = greater oppressive leadership

Table A2: Proposed Questionnaire Items for the Outside Mind Leadership Model

Code	Dimension and questionnaire statement	Response
ECI1	I regularly use AI or algorithmic systems to inform leadership decisions.	1-5
ECI2	External technological systems influence how I interpret organisational problems.	1-5
ECI3	AI-generated information influences the alternatives I consider when making decisions.	1-5
ECI4	I routinely incorporate advice from external experts, information systems, or AI into my judgement.	1-5
ECI5	Some important leadership decisions would be difficult for me to make without external cognitive resources.	1-5
CA1	AI helps me consider perspectives or alternatives that I might otherwise overlook.	1-5
CA2	External cognitive resources improve the quality of information available for my decisions.	1-5
CA3	AI helps me analyse complex organisational problems more effectively.	1-5
CA4	External recommendations strengthen rather than replace my	1-5

	professional judgement.	
CA5	Using AI enables me to make more informed leadership decisions while retaining independent judgement.	1-5
CD1	I am reluctant to make important decisions when AI-supported information is unavailable.	1-5
CD2	I usually accept AI recommendations when they conflict with my initial judgement.	1-5
CD3	I increasingly depend on technological systems to determine the preferred course of action.	1-5
CD4	I find it difficult to override recommendations produced by systems perceived as highly accurate.	1-5
CD5	My confidence in making decisions independently has decreased as my reliance on AI has increased.	1-5
DMA1	I critically evaluate AI-generated recommendations before acting upon them.	1-5
DMA2	I seek additional evidence when an AI recommendation appears questionable.	1-5
DMA3	I am willing to reject an AI recommendation when my professional judgement indicates that it is inappropriate.	1-5
DMA4	I can explain why I accepted or rejected an externally generated recommendation.	1-5
DMA5	I retain final independent judgement even when AI strongly recommends a particular action.	1-5
DO1	I consider myself responsible for decisions I make with AI assistance.	1-5
DO2	I accept responsibility for the consequences of recommendations that I choose to implement.	1-5
DO3	Using AI does not reduce my responsibility for the final decision.	1-5
DO4	I regard AI-supported decisions as my decisions when I possess the authority to approve them.	1-5
MA1	I consider the ethical consequences for employees before implementing AI-supported decisions.	1-5
MA2	Human welfare remains important even when an AI recommendation promises greater efficiency.	1-5
MA3	I consider whether AI-supported decisions respect employees' dignity and rights.	1-5
MA4	I would reject an efficient AI recommendation if I believed it would unfairly harm employees.	1-5
MD1	Responsibility for an adverse outcome rests mainly with the algorithm when I follow its recommendation.	1-5
MD2	I feel less personally responsible for decisions generated primarily by an automated system.	1-5
MD3	When organisational policy requires an AI system, responsibility for its decisions belongs mainly to the organisation.	1-5

MD4	Following a technologically generated recommendation reduces my personal responsibility for its consequences.	1-5
MD5	I experience less moral concern about an adverse decision when the recommendation originated from an AI system.	1-5
OP1	AI systems increase management's capacity to monitor employee behaviour.	1-5
OP2	Algorithmic systems increase management's ability to control how work is performed.	1-5
OP3	AI-supported assessments substantially influence opportunities available to employees.	1-5
OP4	Technological systems increase leadership's capacity to make consequential decisions about employees.	1-5
OP5	Employees have less influence over decisions when management relies heavily on algorithmic recommendations.	1-5
EVC1	Employees can question AI-supported decisions that affect them.	1-5
EVC2	Employees receive meaningful explanations for consequential AI-mediated decisions.	1-5
EVC3	Employees can provide contextual information before important AI-supported decisions are finalised.	1-5
EVC4	Employees have access to an effective appeal process for AI-mediated decisions.	1-5
EVC5	Employee concerns can result in an AI-supported decision being reconsidered.	1-5
EADFP1	AI-supported leadership preserves employees' reasonable discretion over their work.	1-5
EADFP2	Employees are treated with dignity when AI systems contribute to decisions affecting them.	1-5
EADFP3	AI-supported decisions are applied fairly across employees.	1-5
EADFP4	Employees are meaningfully involved in decisions concerning technologies that affect their work.	1-5
EADFP5	AI-supported management respects individual circumstances rather than relying solely on standardised classifications.	1-5
EADFP6	Employees retain meaningful human interaction when consequential workplace decisions are made.	1-5
AGA1	My organisation clearly identifies who is accountable for AI-supported decisions.	1-5
AGA2	Important AI recommendations are subject to meaningful human review.	1-5
AGA3	Leaders are expected to justify consequential decisions even when AI supports them.	1-5
AGA4	Employees are informed when AI materially contributes to decisions affecting them.	1-5
AGA5	My organisation has procedures for identifying and correcting unfair AI-	1-5

	supported decisions.	
ERL1	I use AI to improve decisions without surrendering responsibility for them.	1-5
ERL2	Technology helps me make decisions that better consider employees' needs and circumstances.	1-5
ERL3	I encourage employees to question AI-supported recommendations when they appear inappropriate.	1-5
ERL4	AI-supported decision-making in my organisation is balanced by meaningful human judgement.	1-5
ERL5	External cognitive resources strengthen my leadership without reducing employee agency.	1-5
OL1	AI-supported monitoring substantially restricts employees' autonomy.	1-5
OL2	Employees are expected to accept algorithmically supported decisions without meaningful challenge.	1-5
OL3	Technological systems are sometimes used to justify decisions that employees cannot meaningfully contest.	1-5
OL4	AI-supported management can reduce employees' opportunities to participate in decisions affecting them.	1-5
OL5	Efficiency objectives sometimes take priority over employee dignity or fairness when AI is used.	1-5
OL6	Heavy reliance on algorithmic systems can concentrate organisational power in management.	1-5

Table A3: Proposed Scoring and Measurement Framework

Construct	Items	Score calculation	Possible mean	Interpretation
External Cognitive Influence	ECI1-ECI5	Sum/5	1-5	Higher = stronger outside-mind incorporation
Cognitive Augmentation	CA1-CA5	Sum/5	1-5	Higher = stronger augmentation
Cognitive Dependency	CD1-CD5	Sum/5	1-5	Higher = stronger dependency
Decision-Making Autonomy	DMA1-DMA5	Sum/5	1-5	Higher = stronger independent scrutiny
Decision Ownership	DO1-DO4	Sum/4	1-5	Higher = stronger responsibility ownership
Moral Agency	MA1-MA4	Sum/4	1-5	Higher = stronger ethical agency
Moral Disengagement	MD1-MD5	Sum/5	1-5	Higher = greater moral distancing/disengagement
Organisational Power	OP1-OP5	Sum/5	1-5	Higher = greater AI-mediated managerial power

Employee Voice/ Contestability	EVC1- EVC5	Sum/5	1-5	Higher = greater voice and ability to challenge
Autonomy/Dignity/Fairness/ Participation	EADFP1- EADFP6	Sum/6	1-5	Higher = stronger employee agency/protection
AI Governance/Accountability	AGA1- AGA5	Sum/5	1-5	Higher = stronger governance safeguards
Empowering/Responsible Leadership	ERL1- ERL5	Sum/5	1-5	Higher = more responsible leadership
Oppressive Leadership	OL1-OL6	Sum/6	1-5	Higher = greater oppressive leadership

Proposed interpretation of scores

For descriptive purposes only, subscale means could provisionally be interpreted as **1.00-1.80 = very low**, **1.81-2.60 = low**, **2.61-3.40 = moderate**, **3.41-4.20 = high**, and **4.21-5.00 = very high**. These cut-points should **not be treated as validated clinical or organisational thresholds**; empirical validation should determine whether categorical interpretation is justified. For inferential research, retaining the subscales as continuous latent or composite measures would generally be preferable. The instrument should also avoid calculating one undifferentiated “Outside Mind score,” because augmentation and dependency are theoretically distinct and may coexist.

Table A4: Proposed Measurement Model Corresponding to Figure 1

Conceptual pathway	Predictor/ domain	Proposed mechanism	Outcome/ domain	Expected theoretical relationship
1	External Cognitive Influence	External cognitive incorporation	Cognitive Augmentation	Greater outside-mind integration may enhance cognition
2	External Cognitive Influence	Reliance/deference	Cognitive Dependency	Greater uncritical reliance may increase dependency
3	Cognitive Augmentation	Critical evaluation	Decision-Making Autonomy	Augmentation may preserve/enhance autonomy
4	Cognitive Dependency	Reduced independent scrutiny	Decision-Making Autonomy	Dependency may diminish autonomy
5	Decision-Making	Decision ownership	Moral Agency	Retained autonomy may

	Autonomy			strengthen moral ownership
6	Cognitive Dependency	Responsibility displacement	Moral Disengagement	Dependency may increase moral distancing
7	Moral Disengagement	Reduced personal accountability	Organisational Power	May facilitate less constrained exercise of power
8	Organisational Power	Control/surveillance/decision authority	Employee Agency	Excessive power may diminish voice and autonomy
9	Employee Voice/Contestability	Challenge and participation	Leadership Outcome	Greater voice may favour responsible leadership
10	AI Governance/Accountability	Protective conditions	Dependency → Oppression pathway	Expected to constrain harmful pathways
11	Combined constructs	Cognitive and moral mechanisms	Responsible Leadership	Augmentation + accountability may favour empowerment
12	Combined constructs	Dependency + disengagement + concentrated power	Oppressive Leadership	Combination may increase oppressive outcomes