Contemporary Research Analysis Journal
e-ISSN: 3050-5909
p-ISSN: 3050-5895
Open Access
Volume 03 Issue 09 September 2026
CrossRef DOI: 10.55677/CRAJ/02-2026-Vol03I09
Page no: 592-603

Human–AI Leadership and Workforce Outcomes in Financial Services: A Structured Evidence Review and Work Design Capability Framework

Golley, Israel Tega[1*], Mgbataogu, Ifeanyi Sunday (Ph.D.)[2], AMEH, Jackson Owoicho.[3]
[1,3] Department of Economics and Management, University of Port Harcourt
[1] https://orcid.org/0009-0008-2609-8612
[2] Department of Finance and Banking, University of Port Harcourt

ABSTRACT: Artificial intelligence (AI) is changing financial-services tasks faster than institutions can redesign jobs, skills, and accountability. Exposure estimates are often misread as forecasts of job loss, while productivity claims neglect leadership and job-quality conditions. This study combines a structured integrative review and directed qualitative content analysis of 18 authoritative institutional documents published from 2019 through 2025 with descriptive synthesis of public AI-exposure evidence. Eight organizational mechanisms were coded: strategic alignment, task redesign, employee participation, learning and skills, human oversight, workforce fairness and well-being, performance measurement, and social dialogue. Task redesign, learning and skills, and performance measurement each appeared in 89% of documents. Strategic alignment appeared in 72%, human oversight in 61%, and employee participation and workforce fairness in 50% each. Exposure evidence consistently indicates that augmentation and transformation are distinct from technological feasibility of automation; outcome data remain insufficient to infer net financial-sector employment effects. The proposed Human–AI Work Design Capability framework explains how leaders convert AI exposure into four pathways: task complementarity, accountable judgment, learning velocity, and distributive legitimacy. Sustainable performance requires redesigning bundles of tasks, protecting meaningful human authority, measuring quality and risk alongside speed, and sharing transition benefits. The paper provides testable propositions, a managerial scorecard, and a reproducible research agenda.

KEYWORDS: human AI collaboration, leadership, financial services, future of work, job quality, generative artificial intelligence, organizational capability, workforce transformation.

Cite the Article: *Golley, I.T., Mgbataogu, I.S., AMEH, J. O. (2026). Human–AI Leadership and Workforce Outcomes in Financial Services: A Structured Evidence Review and Work Design Capability Framework. Contemporary Research Analysis Journal, 3(9), 592–603. https://doi.org/10.55677/CRAJ/02-2026-Vol03I09*

Publication Date: September 19, 2026

**Corresponding Author: Golley, Israel Tega*

1. INTRODUCTION

1.1 Background and Context

Artificial intelligence has entered the everyday work of banking, insurance, investment management, payments, and financial regulation. Applications include fraud detection, anti-money-laundering review, credit analysis, customer service, claims processing, software development, research, document summarization, compliance, and personalized advice. Generative AI extends automation from structured prediction into language-intensive work previously treated as the domain of professional judgment. The organizational question is no longer whether financial employees will encounter AI, but how work will be allocated, supervised, evaluated, and rewarded when human and machine contributions are interdependent.

Financial services is an especially consequential setting. The sector contains many information-intensive occupations and standardized workflows, making task exposure comparatively high. At the same time, decisions can create legal, prudential, fiduciary, and consumer consequences. A generated answer that saves minutes can also introduce factual error, privacy leakage, discrimination, unsuitable advice, or misplaced accountability. Productivity cannot therefore be assessed through output volume alone. Institutions must consider decision quality, conduct risk, operational resilience, customer outcomes, workforce capability, and the distribution of gains and burdens.

Public debate often collapses three different concepts. Exposure means that AI can affect some tasks in an occupation. Automation means technology substitutes for human activity. Augmentation means technology increases the productivity or quality of human

activity. Net employment additionally depends on demand, new tasks, organizational choices, prices, regulation, and adjustment. The International Labour Organization (ILO) finds that transformation of jobs is more likely of generative AI than wholesale occupational disappearance, while emphasizing unequal exposure and job-quality risks (Gmyrek et al., 2023, 2025). The International Monetary Fund (IMF) estimates that about 40% of global employment is exposed to AI, with higher exposure in advanced economies, but distinguishes occupations with high complementarity from those facing substitution risk (Cazzaniga et al., 2024). Neither estimate is a prediction that the corresponding share of jobs will vanish.

Leadership shapes the conversion. Executives choose use cases and investment priorities; managers redesign workflows and performance systems; risk leaders define permissible reliance; human-resource leaders organize learning and mobility; and frontline employees discover failure modes that centralized teams cannot anticipate. A technically capable model may produce little value if workers do not trust it, cannot challenge it, or lack time to learn. Conversely, enthusiastic adoption may create hidden work, surveillance, deskilling, or unsafe overreliance. Leadership is therefore an operating capability that connects technology to work design.

1.2 Research Problem and Gap

Three gaps motivate this study. First, exposure studies measure technical possibility but are regularly interpreted as realized workforce outcomes. Occupations are task bundles, and the same technology may automate documentation while augmenting investigation or relationship management. Without task-level implementation data, claims about layoffs, wages, or productivity exceed the evidence.

Second, AI adoption research often treats leadership as executive sponsorship. Sponsorship matters, but human–AI work requires more specific practices: task decomposition, role clarity, employee participation, escalation authority, learning time, performance metrics, and transition support. The mechanisms connecting leadership to workforce outcomes remain fragmented across information systems, organizational behavior, labor economics, human factors, and regulatory guidance.

Third, performance and job quality are analyzed separately. An institution may report faster handling time while rework, error escalation, employee strain, customer dissatisfaction, or long-run deskilling increases. Sustainable performance is multi-dimensional. It includes productivity, quality, innovation, learning, well-being, fairness, retention, and responsible customer outcomes.

1.3 Objectives, Research Questions, and Propositions

The study aims to identify convergent leadership and work-design mechanisms, clarify the interpretation of exposure evidence, develop a testable organizational framework, and provide measures for future causal research.

RQ1: Which leadership and workforce mechanisms recur across authoritative evidence on AI-enabled work?

RQ2: How should AI exposure be distinguished from automation, augmentation, and realized employment outcomes in financial services?

RQ3: Through what organizational pathways can human–AI work design influence productivity, judgment, learning, well-being, and distributive legitimacy?

P1: AI capability has a stronger positive association with risk-adjusted performance when leaders redesign task bundles rather than merely deploy tools into existing workflows.

P2: Learning investment and employee participation mediate the relationship between AI adoption and effective use, error detection, and innovation.

P3: Meaningful human authority moderates the relationship between AI reliance and decision quality; nominal review without information, time, competence, or escalation power produces weaker outcomes.

P4: Productivity measurement that incorporates quality, risk, and rework produces more durable gains than measurement focused on speed or labor reduction alone.

P5: Perceived distributive and procedural fairness mediates the relationship between workforce transition practices and employee trust, retention, and discretionary learning.

1.4 Significance of the Study

The article connects financial technology with leadership and organizational management. It contributes theoretically by integrating task-based economics, socio-technical systems theory, dynamic capabilities, job-demands–resources theory, and organizational justice. It contributes methodologically through a transparent evidence matrix and a strict separation of exposure indicators from realized outcomes. It contributes managerially through a work-design framework and balanced scorecard. Finally, it contributes to policy by identifying the data required to evaluate whether AI generates broadly shared productivity or transfers risk and adjustment costs to workers and customers.

2. LITERATURE REVIEW

2.1 Theoretical Framework

The task approach to technological change treats occupations as collections of activities rather than indivisible units. Technology can substitute for labor in some tasks, complement workers in others, and create new tasks that restore labor demand (Acemoglu & Restrepo, 2019). This approach is particularly useful for finance. A credit analyst's role may include gathering information, cleaning data, generating an initial memo, challenging assumptions, communicating with clients, and accepting accountability. Generative AI can affect each task differently. Occupational exposure therefore does not identify the final job design.

Socio-technical systems theory argues that performance emerges from joint optimization of technical and social systems. Installing a tool without modifying workflow, roles, incentives, and feedback can create bottlenecks or workarounds. Human–AI collaboration is not a fixed property of a model. It is designed through interfaces, decision rights, review thresholds, staffing, and organizational norms.

Dynamic-capabilities theory explains how institutions adapt under technological uncertainty (Teece et al., 1997). Leaders must sense relevant capabilities, seize use cases through resource commitments, and reconfigure routines as evidence changes. Learning velocity matters because generative models, regulations, and employee practices evolve quickly. Training cannot be a one-time course; institutions require ongoing experimentation, communities of practice, incident learning, and mobility pathways.

Job-demands–resources theory predicts that technology can simultaneously add resources and demands (Bakker & Demerouti, 2007). AI may reduce repetitive documentation, provide timely knowledge, and expand autonomy. It may also increase monitoring, pace, cognitive vigilance, exception complexity, and insecurity. Outcomes depend on whether resources—control, support, learning, clarity, recovery time—offset demands. This helps explain why identical tools can improve engagement in one team and accelerate burnout in another.

Organizational justice theory adds distributive, procedural, and interpersonal dimensions. Employees evaluate who receives productivity gains, how job changes are decided, whether voice is meaningful, and whether affected people are treated with dignity (Colquitt et al., 2001). A transition perceived as fair can support experimentation and knowledge sharing. A transition framed solely as head-count reduction can encourage concealment, resistance, or superficial compliance.

These lenses support the Human–AI Work Design Capability (HAWDC) framework. Leadership converts AI exposure into outcomes through four pathways: task complementarity, accountable judgment, learning velocity, and distributive legitimacy. Strategic alignment and task redesign configure complementarity; human oversight protects judgment; participation and skills accelerate learning; and fairness, measurement, and social dialogue support legitimacy.

2.2 Review of Prior Work

A broader organizational behavior literature situates this exposure evidence within firm-level practice: Bankins et al. (2024) synthesize empirical studies of human–AI collaboration, algorithmic management, and labor-market effects, concluding that outcomes depend on how organizations design roles and oversight around AI rather than on the technology alone. Exposure research maps AI capabilities to occupational tasks. Felten et al. (2021) constructed an occupational AI exposure measure based on links between AI applications and human abilities. Generative-AI studies extend this logic to language tasks. Gmyrek et al. (2023) estimate potential effects on job quantity and quality and conclude that transformation is more likely than full automation for most occupations. The refined ILO index (Gmyrek et al., 2025) incorporates changes in model capabilities and global occupational structures. Such indices are valuable for identifying where adaptation may be needed, but they do not observe firm adoption, cost, regulation, worker response, demand elasticity, or job creation.

Macroeconomic analysis likewise separates exposure and complementarity. Cazzaniga et al. (2024) estimate that almost 40% of global employment is exposed and approximately 60% in advanced economies. High exposure can produce either benefit or harm depending on complementarity. Distribution matters because high-income workers may be positioned to capture productivity gains, while other workers face displacement or weakened bargaining power. Country preparedness, digital infrastructure, education, and social protection shape outcomes.

Field evidence shows heterogeneous productivity effects. In customer support, Brynjolfsson et al. (2023) found that access to a generative-AI assistant increased productivity on average, with larger gains among less experienced workers. The technology helped diffuse practices associated with stronger performers. However, one organizational setting does not establish effects for regulated financial judgment. Tasks with verifiable answers differ from suitability assessments, credit decisions, or suspicious-activity investigations, where errors may be rare, consequential, and difficult to observe immediately.

Experimental evidence indicates that generative AI can improve speed and quality on some professional writing tasks, while benefits vary by baseline skill and task fit (Noy & Zhang, 2023). Dell'Acqua et al. (2023) describe a "jagged technological frontier": workers benefit on tasks within model capability but may perform worse when they rely on AI outside that frontier. Leadership must therefore define not only access but task boundaries and verification protocols.

Research on algorithmic management highlights a different risk. AI can allocate work, monitor performance, recommend decisions, and shape evaluation, performing management functions that were previously the preserve of human supervisors (Parent-Rocheleau & Parker, 2022). These systems may reduce discretion and intensify work even when they do not eliminate jobs. The ILO and Organisation for Economic Co-operation and Development (OECD) emphasize consultation, transparency, skills, and worker protection. Job quality includes autonomy, security, intensity, voice, fairness, and development—not only employment count.

International Labour Organization (2024) situates these mechanisms within a broader divide in AI readiness, arguing that digital infrastructure, skills, and social-protection systems determine whether transformation benefits workers or widens inequality between and within countries. The Organisation for Economic Co-operation and Development's employment outlook (2023) similarly finds that AI adoption is accelerating but uneven across firms and workers, and calls for social dialogue, retraining, and monitoring of job quality alongside productivity. Earlier OECD guidance on ethical risk in the workplace identified human oversight, transparency, and worker consultation as preconditions for trustworthy deployment (Salvi del Pero et al., 2022).

Technical and financial-stability standards add a complementary governance layer that the workforce literature rarely engages directly. The National Institute of Standards and Technology's risk management framework (2023) and its generative-AI profile (2024) operationalize trustworthiness, accountability, and human oversight as design requirements rather than aspirations, while the International Organization for Standardization's AI management-system standard (2023) specifies organizational processes for risk assessment, competence, and continual improvement. At the system level, the Financial Stability Board (2024) cautions that common reliance on a small number of third-party AI and data providers could concentrate operational and vendor risk across the financial system, extending governance concerns from the individual firm to the sector as a whole.

Financial-sector evidence shows rapid adoption combined with governance gaps. The Bank of England and Financial Conduct Authority (2024) reported that 75% of surveyed firms were using AI, up from the share reporting machine-learning use in their earlier survey, and another 10% planned adoption within three years. Firms identified third-party dependency, data, cybersecurity, explainability, and skills among relevant constraints. The U.S. Department of the Treasury (2024) similarly found broadening use and recommended coordination, information sharing, compliance review, and support for smaller firms facing capability gaps. Qualitative evidence from bank managers corroborates these survey findings: Moharrak and Mogaji (2025) report that generative AI is already integrated into customer-facing and back-office banking functions, but that managerial preparedness, regulatory compliance, and data-privacy safeguards lag behind deployment.

Leadership scholarship suggests that effective digital transformation combines direction, enabling structures, experimentation, and sensemaking. Yet charismatic advocacy is insufficient. Managers must translate abstract strategy into workflows and protect upward communication about failures. Psychological safety is important because employees are often first to recognize hallucinations, biased outputs, customer confusion, or unsafe automation. If challenging an AI initiative is interpreted as resistance, institutions suppress valuable risk information.

Human oversight is also frequently under-specified. A person may click approval without understanding model limitations or possessing authority to change the outcome; oversight requirements imposed without the resources to exercise them can create a false sense of protection while leaving accountability for algorithmic harm diffuse (Green, 2022). Meaningful oversight requires competence, relevant information, adequate time, independence, and escalation power. It must be placed where human judgment adds value rather than applied indiscriminately. Excessive manual review can create automation bias through fatigue and may neutralize productivity gains.

2.3 Synthesis of Gaps

The evidence identifies five unresolved issues. First, institutions lack consistent task-level measures linking AI use to output, quality, and workforce change. Second, leadership practices are discussed broadly but rarely operationalized. Third, productivity studies are concentrated in a limited set of tasks and firms. Fourth, employee voice, fairness, and well-being are less consistently connected to financial-risk governance. Fifth, long-run effects on skill formation remain uncertain: AI may accelerate learning through coaching or produce deskilling when workers no longer practice foundational reasoning.

The present study addresses these gaps through mechanism coding and a framework that treats work design as a capability. It does not claim that documentary prevalence demonstrates effectiveness. Instead, it identifies convergence, omissions, and propositions suitable for causal testing.

3. METHODOLOGY

3.1 Research Design

The study uses a structured integrative review, directed qualitative content analysis, and descriptive synthesis of public exposure and adoption evidence. The design fits a multidisciplinary question involving economics, management, labor standards, financial regulation, and organizational practice. It is not presented as an exhaustive systematic review or meta-analysis because outcomes, units, and methods are heterogeneous.

The unit of qualitative analysis is one authoritative document. The analytical sequence included question definition, source search, eligibility screening, codebook development, document-level coding, frequency and co-occurrence analysis, negative-case examination, and theory construction. Exposure statistics were interpreted according to their original definitions and were not converted into job-loss forecasts.

3.2 Population and Sampling

The conceptual population includes English-language public reports, standards, surveys, and guidance addressing AI, work design, skills, employment, human oversight, or financial-services adoption. Purposive maximum-variation sampling covered international labor institutions, economic organizations, standards bodies, national financial authorities, and selected foundational evidence.

Documents were eligible when they (a) were issued by an authoritative public, intergovernmental, or standards institution; (b) contained empirical exposure/adoption evidence or substantive organizational recommendations; (c) were available in full; and (d) were published from 2019 through 2025, except foundational standards retained for theoretical continuity. Promotional reports without transparent methods, short news summaries duplicating a full report, and sources making unsupported employment forecasts were excluded.

Eighteen documents were retained. The sample intentionally combines cross-sector workforce sources and financial-sector sources. Cross-sector evidence supplies labor and work-design concepts, while finance-specific evidence supplies institutional context. Frequency therefore represents emphasis in this designed corpus, not prevalence in all publications.

3.3 Data Collection and Variables

For each document, the study recorded issuer, title, year, scope, and URL. Eight binary mechanisms were specified:

- **SA—Strategic alignment:** linkage of AI adoption to business purpose, public value, risk appetite, governance responsibility, or executive direction.
- **TR—Task redesign:** decomposition or reconfiguration of tasks, workflows, jobs, roles, staffing, or human–machine allocation.
- **EP—Employee participation:** consultation, co-design, worker voice, feedback, psychological safety, or participation in implementation.
- **LS—Learning and skills:** training, reskilling, AI literacy, managerial capability, professional development, or mobility pathways.
- **HO—Human oversight:** human review, contestability, judgment, accountability, escalation authority, or limits on automated decisions.
- **WF—Workforce fairness and well-being:** job quality, inclusion, work intensity, privacy, equality, security, or distributive consequences.
- **PM—Performance measurement:** evaluation of productivity, quality, risk, errors, workforce outcomes, or longitudinal impact.
- **SD—Social dialogue and transition governance:** collective consultation, employer–worker coordination, public-private partnership, adjustment policy, or social protection.

A code received 1 only when the document offered a substantive analysis, recommendation, or evidence relevant to the mechanism. Incidental mention received 0. Multiple codes were allowed. Ambiguity was resolved conservatively.

The descriptive evidence table reports selected published estimates of exposure or adoption. Measures retain their original populations and should not be compared as if they share a denominator. “Exposure,” “use,” and “planned use” are distinct variables. No pooled effect was calculated.

3.4 Data Analysis

Mechanism prevalence was calculated as the number of coded documents divided by 18. Pairwise co-occurrence identified bundles, while negative cases identified documents emphasizing productivity without workforce participation or job quality. No inferential significance tests were used because sampling was purposive.

Pattern matching mapped mechanisms to the four HAWDC pathways. Task complementarity was associated principally with SA and TR. Accountable judgment was associated with HO and PM. Learning velocity was associated with EP and LS. Distributive legitimacy was associated with WF, EP, and SD. These mappings are theoretical interpretations, not factor-analysis results.

The public-data synthesis applied three validity rules. First, exposure was not described as displacement. Second, survey adoption was not described as effective deployment. Third, productivity evidence from one task or organization was not generalized to the entire financial sector. These rules protect against common category errors.

3.5 Validity, Reliability, and Research Integrity

Construct validity was supported through explicit operational definitions and established theories. Source triangulation included labor, economic, technical, and financial authorities. An audit trail appears in Appendix A and the accompanying CSV. Negative-case analysis prevented the framework from assuming all adoption is participatory or beneficial.

Reliability is limited by single-author coding. The matrix was coded in two passes after rechecking definitions, with uncertain cases coded 0. Future research should employ independent coders and report Cohen's kappa or Krippendorff's alpha. Passage-level quotations and archival source versions should be preserved in a registered repository where copyright permits.

Internal validity is limited because documentary emphasis cannot establish causal relationships. External validity is bounded by public English-language evidence and highly regulated contexts. Estimates of occupational exposure depend on task taxonomies and capability assumptions. Adoption surveys may reflect response and self-report bias.

4. RESULTS

4.1 Public Evidence on Exposure, Adoption, and Outcomes

Table 1. Selected Public Indicators Relevant to Human–AI Work

Source and population	Indicator	Estimate	Correct interpretation
IMF, global employment	Jobs exposed to AI	About 40%	Tasks may be affected; exposure is not displacement.
IMF, advanced economies	Jobs exposed to AI	About 60%	Higher cognitive-task exposure includes complementary roles.
ILO, global occupations	Central expected effect	Transformation	Most occupations contain mixed automatable and human tasks.
Bank of England/FCA, surveyed financial firms, 2024	Firms using AI	75%	Organizational adoption; not employee-level use or effectiveness.
Bank of England/FCA, surveyed firms, 2024	Planning use within three years	10%	Stated intention; not realized deployment.

Note. Indicators retain different definitions and denominators and must not be pooled. IMF values are approximate exposure estimates. ILO findings concern potential task effects. Bank of England/FCA values describe responding firms.

The evidence supports high potential exposure in advanced, information-intensive economies, but not a numerical forecast of financial job losses. Exposure includes complementarity. A financial professional may use AI for retrieval and drafting while retaining responsibility for judgment, communication, and exception handling. Net employment further depends on demand expansion, new products, costs, regulation, and organizational design.

The 75% firm-level adoption estimate indicates that workforce governance is already a practical concern rather than a distant scenario. Yet institution-level use can mean anything from a limited fraud model to widespread generative assistants. Without task-level usage, outcome, and workforce data, adoption rates cannot establish productivity or job quality.

4.2 Prevalence of Leadership and Workforce Mechanisms

Table 2. Mechanism Prevalence Across the 18-Document Corpus

Mechanism	Documents	Prevalence	Interpretation
Task redesign (TR)	16	88.9%	Outcomes depend on reconfiguring work rather than adding a tool.
Learning and skills (LS)	16	88.9%	Capability building is a consistent transition response.

Mechanism	Documents	Prevalence	Interpretation
Performance measurement (PM)	16	88.9%	Evaluation is common, although balanced workforce measures remain uneven.
Strategic alignment (SA)	13	72.2%	AI use is expected to connect to purpose, risk, and governance.
Human oversight (HO)	11	61.1%	Judgment receives less consistent attention than measurement.
Employee participation (EP)	9	50.0%	Worker knowledge is recognized, but participation is uneven.
Workforce fairness/well-being (WF)	9	50.0%	Job quality and distribution appear less often than skills or oversight.
Social dialogue/transition governance (SD)	8	44.4%	Collective adjustment receives the least consistent attention.

Learning and skills dominated the corpus. This convergence is useful but incomplete. Training cannot correct a poorly designed workflow or remove conflicting incentives. Skill programs create value when employees can apply learning, receive feedback, and move into redesigned roles.

Human oversight appeared in eleven documents. However, documents varied in specificity. Strong formulations linked oversight to authority, accountability, transparency, and contestability. Weak formulations merely required a person somewhere in the process. The distinction supports P3: oversight quality depends on organizational conditions.

Task redesign appeared in sixteen documents and strategic alignment in thirteen, indicating that adoption is treated as organizational change rather than software procurement. Although sixteen documents included some form of performance measurement, only nine addressed fairness or well-being; balanced workforce measurement remained less developed than technical or productivity evaluation.

4.3 Mechanism Bundles

Learning and skills co-occurred with task redesign in fifteen documents. This is the most important implementation bundle: training content should follow the future task architecture, while task design should reflect learnability and career progression. Generic prompt training without role redesign may increase isolated use but not organizational capability.

Strategic alignment and human oversight co-occurred in ten documents. This bundle links executive intent to decision accountability. Risk appetite becomes operational only when teams know which tasks may be delegated, what evidence must be reviewed, and who can stop deployment.

Employee participation and workforce fairness co-occurred in six documents. Participation appears to be a mechanism for both knowledge and legitimacy. Frontline staff reveal undocumented work, customer consequences, accessibility barriers, and exception patterns. Consultation after design choices are fixed cannot provide the same informational benefit as co-design.

Performance measurement co-occurred with task redesign in fourteen documents. The negative cases are nevertheless revealing: measurement commonly emphasized exposure, productivity, or technical risk rather than a balanced account of workforce experience and distribution. Leaders may consequently measure adoption counts or processing speed without evaluating error, rework, escalation, customer harm, learning, or work intensity.

Across the coding matrix, the highest pairwise co-occurrence was between task redesign and learning and skills (15/18 documents, 83.3%), followed by task redesign and performance measurement (14/18, 77.8%). These counts indicate recurring documentary bundles, not statistical associations or evidence that one mechanism causes another.

4.4 Human–AI Work Design Capability Framework

The evidence supports four conversion pathways.

Task complementarity begins with task decomposition and allocates activity according to comparative strengths. AI may provide speed, pattern recognition, or drafting; humans provide contextual judgment, ethical responsibility, relationship management, and novel exception handling.

Accountable judgment ensures that decision rights, review evidence, and escalation authority match consequence.

Learning velocity refers to the organization's ability to develop skills and update practice faster than tools and risks change. It depends on protected learning time, peer communities, experimentation, feedback, and mobility.

Distributive legitimacy concerns whether workers and customers view decisions and benefit allocation as fair. It depends on voice, transparency, transition support, privacy, accessibility, and credible sharing of productivity gains.

These pathways generate a feedback loop. Better task design produces more informative performance data. Meaningful oversight identifies errors and new training needs. Participation reveals hidden demands. Fair transition practices increase trust and willingness to report problems. Conversely, labor-reduction targets can suppress learning and encourage premature automation.

4.5 Outcome Architecture

The review indicates that no single productivity metric is adequate. Institutions should measure at least five outcome families: operational performance, decision quality, customer and risk outcomes, learning and innovation, and workforce outcomes. Speed should be paired with accuracy and rework. Cost should be paired with incidents and customer remedy. Use should be paired with validated competence. Employee sentiment should be paired with turnover, mobility, workload, and access to development.

Attribution requires appropriate comparison. Before–after changes may reflect demand or policy shifts. Stronger designs include staggered rollout, matched teams, randomized access where ethical, interrupted time series, and difference-in-differences. Analyses should report heterogeneous effects by experience, role, gender, race or ethnicity where legally and ethically appropriate, disability, contract status, and location. Aggregate gains can conceal concentrated harm.

5. DISCUSSION

5.1 Interpretation of Results

The findings show broad consensus around skills, oversight, strategy, and task redesign, but weaker integration of measurement, fairness, participation, and social dialogue. This pattern describes an implementation risk: institutions may invest in tools and training while leaving jobs, incentives, and transition arrangements substantially unchanged.

High exposure in finance should be understood as managerial responsibility, not predetermined displacement. Leaders choose whether AI removes low-value work, expands service demand, intensifies monitoring, concentrates expertise, or creates development pathways. Technology sets a frontier of possibilities; organization design influences the realized point.

Skills are necessary but should not become an individualized response to structural change. Employees cannot “reskill” their way out of unclear accountability, unrealistic workloads, or disappearing entry-level learning tasks. Institutions must preserve pathways by which junior workers acquire judgment. If AI performs the drafts and routine analyses through which novices formerly learned, leaders need deliberate apprenticeships, simulations, rotation, and review.

5.2 Comparison with Prior Literature

The results align with the task approach: occupations contain substitutable and complementary activities. They also align with the jagged-frontier evidence, which shows that task fit matters. The HAWDC framework extends these insights by specifying organizational conversion mechanisms.

Socio-technical theory is supported by the high prevalence of task redesign. Performance cannot be inferred from model capability because workflows redistribute work. AI can remove documentation yet increase verification and exception handling. Hidden human labor must be measured.

Job-demands–resources theory explains heterogeneous workforce outcomes. A tool can provide cognitive support while increasing pace and vigilance. Leaders influence the balance through staffing, autonomy, learning time, interface quality, and recovery. Organizational justice explains why participation and transition support affect trust beyond instrumental effectiveness.

Dynamic-capabilities theory helps interpret learning as an institutional process. Rapidly changing models make static skills inventories obsolete. Firms need mechanisms to sense new capabilities, test them, reconfigure roles, and retire unsafe practices. The relevant advantage is not possession of a common model but the capacity to learn responsibly.

5.3 Theoretical Implications

The framework makes three theoretical contributions. First, it treats leadership as a meso-level conversion capability between occupational exposure and realized outcomes. Second, it integrates job quality into performance rather than treating workforce impacts as an external constraint. Third, it identifies meaningful human authority as a scarce organizational resource. Oversight depends on time, skill, information, independence, and power.

The five propositions can be tested using multilevel models in which employees are nested in teams and institutions. Task-level AI use and work design would predict quality, productivity, learning, and well-being. Mediation analysis could examine participation and fairness, while longitudinal designs could identify adaptation. A configurational method could test whether different bundles produce comparable performance.

5.4 Managerial and Policy Implications

Leaders should begin with tasks, not job titles. Each workflow should identify purpose, inputs, AI contribution, human contribution, failure severity, verification, escalation, and learning value. High-consequence decisions require stronger evidence and protected human authority. Low-risk administrative work can use lighter controls.

Participation should occur before procurement and workflow lock-in. Cross-functional design teams should include frontline staff, risk, compliance, technology, human resources, accessibility expertise, and customer representatives where appropriate. Employees need protected channels to report failure without being labeled resistant.

Training should be role-based and continuous. Baseline AI literacy is insufficient for validators, managers, advisors, investigators, and developers. Learning time should be budgeted as work, and assessment should use realistic tasks. Institutions should monitor whether junior roles lose developmental content and create alternative pathways.

Performance systems should resist easy but misleading measures. Output volume and labor hours may improve while rework or risk migrates downstream. Balanced evaluation should include quality, exceptions, customer outcomes, control effectiveness, employee strain, skill development, and distribution. Incentives should not reward bypassing review.

Transition governance should specify redeployment, internal mobility, reskilling access, notice, consultation, and support. Shared gains can include wage progression, reduced low-value work, better scheduling, learning, or reinvestment in service quality. Policy makers should improve task-level data, portable training support, worker protection, and consultation mechanisms while avoiding claims that exposure estimates are forecasts.

Table 3. Human–AI Leadership Scorecard

Domain	Leading indicator	Outcome indicator	Leadership question
Strategy	Use cases with defined value and risk thesis	Risk-adjusted benefit realized	What problem is AI solving?
Task design	Workflows decomposed and reviewed	Time, quality, rework, exceptions	Are tasks allocated by comparative strength?
Participation	Frontline roles represented in design	Adopted employee improvements	Is voice early and consequential?
Learning	Protected hours and role-based assessment	Competence, mobility, error detection	Can capability evolve with the tool?
Oversight	Decisions with named accountable owner	Overrides, escapes, customer outcomes	Can humans understand and change outcomes?
Well-being	Workload and surveillance impact reviewed	Strain, autonomy, retention, absence	Are resources keeping pace with demands?
Fairness	Transition and access analysis completed	Distribution of gains and harms	Who benefits, who bears risk, and why?
Measurement	Baselines and comparison designs defined	Durable multi-dimensional performance	Are we measuring value rather than activity?

5.5 Limitations

The corpus is purposive, public, English-language, and institutionally weighted. Binary coding compresses nuance. Documentary frequency does not establish implementation or effect. Single-author recoding is not independent reliability testing. Exposure estimates rely on modeled task-capability correspondence rather than observed job change. Adoption data are firm-reported and do not identify intensity or employee coverage.

The framework has not been tested with employee-level or firm-level data. Causal relationships remain propositions. Financial-services subsectors differ substantially, and national labor institutions affect transferability. Rapid technology change may alter task capabilities after publication. These limitations favor transparent updates rather than false precision.

6. CONCLUSION AND RECOMMENDATIONS

6.1 Summary of Key Findings

Human–AI outcomes in financial services are not determined by exposure alone. Across 18 authoritative documents, task redesign, learning and skills, and performance measurement were the most prevalent mechanisms, followed by strategic alignment. Employee participation, workforce fairness, and social dialogue appeared less consistently. This gap matters because technical deployment can increase output while degrading judgment, learning, job quality, or legitimacy.

The HAWDC framework explains how leadership converts AI exposure through task complementarity, accountable judgment, learning velocity, and distributive legitimacy. It treats responsible workforce design as a productive capability. The central implication is that organizations should automate or augment tasks only after defining the human contribution, decision authority, learning pathway, outcome measures, and transition obligations.

6.2 Recommendations for Research and Practice

Financial leaders should map task bundles, establish role-specific human authority, involve employees early, protect continuous learning, and measure quality, risk, rework, customer outcomes, well-being, and distribution alongside productivity. Regulators and labor institutions should promote task-level reporting and distinguish exposure from realized automation. Educational institutions should combine AI literacy with domain judgment, ethics, communication, and verification.

Future studies should use longitudinal and quasi-experimental designs across financial subsectors. Minimum measures should include task-level use, time, quality, exceptions, errors, customer outcomes, worker autonomy, work intensity, skill acquisition, mobility, pay, and retention. Only evidence that integrates economic performance and human outcomes can determine whether AI-enabled finance is genuinely productive and socially sustainable.

REFERENCES

1. Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives, 33*(2), 3–30. https://doi.org/10.1257/jep.33.2.3
2. Bakker, A. B., & Demerouti, E. (2007). The job demands-resources model: State of the art. *Journal of Managerial Psychology, 22*(3), 309–328. https://doi.org/10.1108/02683940710733115
3. Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2024). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of Organizational Behavior, 45*(2), 159–182. https://doi.org/10.1002/job.2735
4. Bank of England & Financial Conduct Authority. (2024). *Artificial intelligence in UK financial services—2024*. https://www.bankofengland.co.uk/report/2024/artificial-intelligence-in-uk-financial-services-2024
5. Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work* (NBER Working Paper No. 31161). National Bureau of Economic Research. https://doi.org/10.3386/w31161
6. Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A. J., Pizzinelli, C., Rockall, E. J., & Tavares, M. M. (2024). *Gen-AI: Artificial intelligence and the future of work* (IMF Staff Discussion Note 2024/001). International Monetary Fund. https://doi.org/10.5089/9798400262548.006
7. Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology, 86*(3), 425–445. https://doi.org/10.1037/0021-9010.86.3.425
8. Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). *Navigating the jagged technological frontier* (Harvard Business School Working Paper No. 24-013). https://www.hbs.edu/faculty/Pages/item.aspx?num=64700
9. Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal, 42*(12), 2195–2217. https://doi.org/10.1002/smj.3286
10. Financial Stability Board. (2024). *The financial stability implications of artificial intelligence*. https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/
11. Gmyrek, P., Berg, J., & Bescond, D. (2023). *Generative AI and jobs: A global analysis of potential effects on job quantity and quality* (ILO Working Paper No. 96). International Labour Organization. https://doi.org/10.54394/FHEM8239
12. Gmyrek, P., Berg, J., Kamiński, K., Konopczyński, F., Ładna, A., Nafradi, B., Rosłaniec, K., & Troszyński, M. (2025). *Generative AI and jobs: A refined global index of occupational exposure* (ILO Working Paper No. 140). International Labour Organization. https://doi.org/10.54394/HETP0387

13. Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review, 45*, Article 105681. https://doi.org/10.1016/j.clsr.2022.105681
14. International Labour Organization. (2024). *Mind the AI divide: Shaping a global perspective on the future of work*. https://www.ilo.org/publications/major-publications/mind-ai-divide
15. International Organization for Standardization. (2023). *ISO/IEC 42001:2023 information technology—Artificial intelligence—Management system*. https://www.iso.org/standard/81230.html
16. Moharrak, M., & Mogaji, E. (2025). Generative AI in banking: Empirical insights on integration, challenges and opportunities in a regulated industry. *International Journal of Bank Marketing, 43*(4), 871–896. https://doi.org/10.1108/IJBM-08-2024-0490
17. National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1
18. National Institute of Standards and Technology. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile* (NIST AI 600-1). https://doi.org/10.6028/NIST.AI.600-1
19. Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192. https://doi.org/10.1126/science.adh2586
20. Organisation for Economic Co-operation and Development. (2023). *OECD employment outlook 2023: Artificial intelligence and the labour market*. https://doi.org/10.1787/08785bba-en
21. Parent-Rocheleau, X., & Parker, S. K. (2022). Algorithms as work designers: How algorithmic management influences the design of jobs. *Human Resource Management Review, 32*(3), Article 100838. https://doi.org/10.1016/j.hrmr.2021.100838
22. Salvi del Pero, A., Wyckoff, P., & Vourc'h, A. (2022). *Using artificial intelligence in the workplace: What are the main ethical risks?* (OECD Social, Employment and Migration Working Papers No. 273). OECD Publishing. https://doi.org/10.1787/840a2d9f-en
23. Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7), 509–533. https://doi.org/10.1002/(SICI)1097-0266(199708)18:7%3C509::AID-SMJ882%3E3.0.CO;2-Z
24. U.S. Department of the Treasury. (2024). *Artificial intelligence in financial services*. https://home.treasury.gov/system/files/136/Artificial-Intelligence-in-Financial-Services.pdf

Appendix A: Document-Level Coding Matrix

ID	Issuer and document (year)	SA	TR	EP	LS	HO	WF	PM	SD
W01	IMF, *Gen-AI and the Future of Work* (2024)	1	1	0	1	0	1	1	1
W02	ILO, *Generative AI and Jobs* (2023)	0	1	1	1	1	1	1	1
W03	ILO, *Refined Global Exposure Index* (2025)	0	1	0	1	0	1	1	1
W04	ILO, *Mind the AI Divide* (2024)	1	1	1	1	1	1	0	1
W05	OECD, *Employment Outlook 2023* (2023)	1	1	1	1	1	1	1	1
W06	OECD, *AI in the Workplace* (2022)	1	1	1	1	1	1	0	1
W07	Bank of England/FCA, *AI in UK Financial Services* (2024)	1	1	0	1	1	0	1	0
W08	U.S. Treasury, *AI in Financial Services* (2024)	1	1	0	1	1	0	1	0
W09	FSB, *Financial Stability Implications of AI* (2024)	1	1	0	1	1	0	1	0
W10	NIST, *AI RMF 1.0* (2023)	1	0	1	1	1	0	1	0
W11	NIST, *Generative AI Profile* (2024)	1	1	1	1	1	1	1	0

ID	Issuer and document (year)	SA	TR	EP	LS	HO	WF	PM	SD
W12	ISO/IEC, *AI Management System* (2023)	1	1	1	1	1	0	1	0
W13	Brynjolfsson et al., *Generative AI at Work* (2023)	0	1	0	1	0	0	1	0
W14	Noy & Zhang, productivity experiment (2023)	0	1	0	1	0	0	1	0
W15	Dell'Acqua et al., *Jagged Frontier* (2023)	1	1	1	1	1	0	1	0
W16	Felten et al., occupational AI exposure (2021)	0	1	0	0	0	0	1	0
W17	Acemoglu & Restrepo, tasks framework (2019)	1	1	0	1	0	1	1	1
W18	Colquitt et al., organizational justice (2001)	1	0	1	0	0	1	1	1

Codes: SA = strategic alignment; TR = task redesign; EP = employee participation; LS = learning and skills; HO = human oversight; WF = workforce fairness and well-being; PM = performance measurement; SD = social dialogue and transition governance. Coding indicates substantive content, not demonstrated effectiveness.

Appendix B: Proposed Task-Level Data Schema

1. **Context:** institution type, business unit, occupation, employment arrangement, jurisdiction, and deployment date.
2. **Task:** standardized task identifier, frequency, duration, consequence, customer impact, and learning value.
3. **AI use:** model and version, voluntary or required use, assistance or automation, usage intensity, and human review.
4. **Operational outcomes:** time, throughput, cost, rework, exceptions, and downtime.
5. **Quality and risk:** accuracy, override, near miss, incident, conduct outcome, complaint, and remediation.
6. **Worker outcomes:** autonomy, cognitive demand, work intensity, well-being, surveillance, satisfaction, and psychological safety.
7. **Capability outcomes:** assessed skill, learning time, knowledge retention, mobility, promotion, and wage progression.
8. **Distribution:** outcome differences by experience, role, demographic group where lawful, disability, location, and contract status.
9. **Transition:** consultation, notice, reskilling offer, redeployment, attrition, redundancy, and social-protection access.

Appendix C: Illustrative Survey and Longitudinal Design

A future study could recruit employees from banking, insurance, payments, investment management, and fintech firms before and after staggered AI deployment. Teams not yet receiving the tool could provide matched comparisons. Pre-registration should specify primary outcomes, exclusions, missing-data rules, and heterogeneous-effect tests.

Example seven-point items include: "I understand which decisions remain my responsibility"; "I have authority to challenge an AI-assisted recommendation"; "AI removes low-value work from my role"; "AI increases the pace or intensity of my work"; "I receive enough time to develop relevant skills"; "employees affected by AI have meaningful input"; "the benefits of AI adoption are shared fairly"; and "I can report an AI-related problem without negative consequences."

Objective measures should include task time, audited quality, rework, escalations, customer outcomes, learning assessment, absenteeism, internal mobility, turnover, and compensation. Researchers should avoid surveillance-heavy collection, obtain informed consent where required, minimize identifiable data, and establish safeguards against employment retaliation. Exposure, use, and outcome measures must remain analytically distinct.