Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

THE AI WORKFORCE PROBLEM: Decision Architecture and Governance Frameworks for Large-Scale Human AI Collaboration in Enterprise Consulting

Karan Dhir

affilication : NA

Location SUNNYVALE CA

CMT mail :karan.dhir@outlook.com

ARTICLE INFO

Received: 10 Oct 2024

Revised: 01 Nov 2024

Accepted: 10 Dec 2024

ABSTRACT

The article explores how artificial intelligence (AI) agents are being integrated on a wide scale into enterprise consulting, with McKinsey & Company reporting that it uses around 25,000 AI agents along with 40,000 human workers for its operations. Based on a systematic literature review of 24 sources and triangulation of secondary case data across 2022-2024, the study explores how AI agents are reshaping the landscape of knowledge work by streamlining the process of knowledge synthesis, data analysis, structured reporting and presentation generation, thereby enhancing the efficiency, throughput, and scalability of tasks. The Quantitative results show a throughput index improvement as high as 171 percentage points as compared to the pre-integration baseline, while conserving the output quality via an iterative governance calibration. The study reveals that sustainable human-AI collaboration is not fundamentally limited by the capabilities of the model or capabilities of the cloud infrastructure deployment, but by the capacity to build robust decision architecture, such as governance procedures and protocols around decisions that involve human-AI collaboration on task allocation, quality assurance, accountability and human oversight. As a conceptual model, it is introduced as a Decision Architecture Framework with five layers, and incorporated in the Methodology Section of the paper. Eight key deployment risks are identified in this alternative Risk Assessment Matrix, based on empirical evidence: Automation complacency, Error attribution gaps and Regulatory non-compliance are among the most critical to mitigate. The human oversight intervention data also show that the bulk of human overrides are occurring in task domains that require judgment, which are more likely the more complex of the domains. It's evident that organizations need to allocate resources simultaneously in the development of AI capabilities and governance architecture to achieve sustainable value from the use of large-scale AI agents.

Keywords: AI Agents, Human–AI Collaboration, Decision Architecture, Enterprise AI Adoption, Organizational Design, Knowledge Work, Automation, AI Governance, Digital Transformation, Automation Risk Management, Scalable AI Systems

1. INTRODUCTION

The rampant expansion and adoption of artificial intelligence (AI) agents in large-scale enterprise environments is one of the most significant shifts in society in the early 21st century. In contrast to earlier automation waves, which tended to automate "rules-based" manual tasks, modern AI agents are breaking into the "knowledge work" itself, tasks that have previously been performed with advanced human cognition, context-sensing and professional

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

judgement (Brynjolfsson & McAfee, 2014; Daugherty & Wilson, 2018). This infusion of AI into these high-value cognitive processes is more than just an efficiency play; it's a paradigm shift in knowledge management, a rethinking of knowledge creation, validation, and dissemination.

The case study by McKinsey & Company, describing the parallel work of some 25,000 AI agents and 40,000 human workers, offers a unique empirical insight into the dynamics of large-scale collaboration between people and AI. Such a situation where AI agents, in purely quantitative terms, are coming close to human workers represents a qualitative leap from previous notions about AI at work, where AI systems operated more like nozzles or pliers, as instruments of labor subservient to human operators. In McKinsey's deployment of AI agents, semi-autonomous agents support complex consulting activities at velocities that were previously unfeasible, such as researching and synthesising information, analysing data, modelling finances, screening for regulatory compliance, and creating presentations.

Nevertheless, this deployment scale, the research literature suggests that the most impactful issues involve such deployments are not technological. The capabilities of the model have progressed to the point where they can power many professional knowledge tasks; cloud infrastructure can provide the ability to run a massive number of agents concurrently; and data pipelines can be created to help the agents have access to relevant and timely information (Chatterjee, 2023; Malaraju & Bondalapati, 2023). Instead, the determining factor is one of organization and architecture: the lack of decision architecture frameworks that could manage whether and how to allocate tasks, how to ensure there is oversight to validate the quality of AI outputs, who will be held accountable if the output turns out wrong, and how a human oversight process might fit in in systems where AI agents are responsible for a large proportion of the organizational output (Kubam, 2024; Kakaraparthi 2022).

This paper aims to answer the questions: (1) How are AI agents changing the way knowledge work is performed in enterprise consulting? (2) What building blocks of decision architecture are needed to scale effective, human-AI collaboration? (3) What are the organizational implications around scaling up AI agents and how can governance structures affect this? The McKinsey case was used in addressing these questions, with literature reviewed using a systematic approach touching on the concepts of AI governance, human factors engineering, organisation design and enterprise AI adoption.

This paper is organized as follows. Theoretical and empirical literature pertinent to the topic is examined (section 2). Section 3 explains the methodological approach used in preparation for this study and introduces the conceptual approach for analysis, a Decision Architecture Framework. The empirical findings based on the quantitative analyses are summarized in Section 4, with deployment patterns resulting from the analyses along with the efficiency improvements, oversight intervention rates and risk assessments. The implications for enterprise AI governance is discussed in section 5. The recommendations and directions for future research are provided in section 6.

2. LITERATURE REVIEW

2.1 Automation and the Economics of Knowledge Work

The basic theoretical ideas of the current research have overlaps with three main fields of literature: (a) the economics and sociology of work automation, (b) human factors and human–AI interaction, and (c) AI governance and organizational design.

The debate about work automation has changed much since the initial discussions of technological unemployment (Acemoglu & Restrepo, 2018). The “task-based” model of automation makes an inherent separation between the tasks that automation substitutes, and new tasks around the automation. In the context of knowledge work, this model suggests that AI agents might replace or augment tasks involving standardized procedures or information (like writing reports with a template, summarizing data), with the added possibility that they may spur demand for higher-level human functions that guide, audit, and interpret AI-generated content. By 2025, around 85 million jobs will presumably fall victim to automation worldwide, the World Economic Forum (2020) estimates, but some 97 million new ones will arise corresponding to the new human-machine division of labour.

One of the most influential theoretical arguments about the cognitive division of labour between humans and AI systems (elligent/automated) is offered by Jarrahi (2018). Building on the notion of "complementarity", Jarrahi

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

suggests we think about AI systems that have extremely good performance in highly structured data processing, pattern recognition and prediction in well-defined conditions whereas humans have even better at areas that demand tacit knowledge, ethical reasoning, contextual thinking and understanding of various interests. The complementarity thesis is highly relevant to enterprise consulting because the kinds of things enterprise consultants can offer clients often have both data-rich, analytical elements (AI-suited) and client-specific, interpretive elements (humansuited).

2.2 Human Factors in Human–AI Collaboration

The dangers of automation have long been documented in the human factors literature and most notably is automation complacency, which refers to a decline in human vigilance and critical engagement when humans are operating in parallel with seemingly reliable automated systems (Parasuraman & Manzey, 2010; Cummings, 2017). Lee and See (2004) determined that judicious levels of trust in automation are necessary for safe and effective human-AI teaming; too much or too little trust leads to less-than-optimal results. In enterprise consulting, such conclusions could be readily, and often carelessly, adopted by human workers who cannot independently evaluate intricate conclusions produced by AI agents tasked with research or financial analysis.

Gombolay et al. (2015) show that the assignment of decision-making power to human or artificial agents in a team has significant impact on team efficiency and human worker satisfaction in an empirical study. Their findings imply that relying on human–AI partnerships that have not been explicitly designed in terms of authority relationships is insufficient for achieving optimal team performance, and that between automation level and human perception of the team's performance is mediated by human perceptions of fairness and control. These insights into human factors are what inspired the explicit human oversight layer included within the Decision Architecture Framework created in this study.

2.3 AI Governance and Organizational Design

Like Dafoe (2018), fairness, human oversight, accountability, and transparency have all been identified as principles for responsible deployment of AI in the literature on AI governances (extended in the ethical literature of Floridi et al., 2018). More recently, Rahwan (2018) has introduced the idea of 'society-in-the-loop' programming to bring the idea of individual behaviour to systemic governance: AI systems should be integrated into social systems that can also recognize, discuss, and correct problematic behaviours in a large scale. In enterprise settings, for example, Amershi et al. (2019) report that failures to use AI systems are often more an issue of mismatch between the capability of the AI system and the organisational structures where it is being used rather than the system's limitations. This is the motivation behind focusing this study on decision architecture.

Shrestha et al, (2019) particularly focus on the impact of AI-supported decision making on the processes within organizations and discuss the risk of "decision offloading", which means over time, as more and more analytical tasks are done by the AI systems, human actors may increasingly lose the ability to engage critically with what has been generated by AI systems. Kamar (2016) proposes a "hybrid intelligence" model in which human agents and artificial agents are regarded as complementary resources in the same task environment and engaged in various tasks based on their respective comparative advantages and capabilities assessment in realtime. While there has been significant theoretical work, literature on the governance of AI agents is lacking at scale; such large numbers of agents running at once and interacting with one another in complex, interdependent business consulting processes are not addressed. In the present study, we aim to help fill in this gap by creating a governance structure that reflects a large scale, real-world deployment.

3. METHODOLOGY

3.1 Research Design Overview

The research design used in this study is mixed methods which combines systematic literature review, single-case study analysis and framework development deductively. The methodological approach can be seen in Figure 1 as three successive stages. Phase 1 (Literature Synthesis) closed with a systematic search of Scopus, Web of Science and Google Scholar using the following search terms and their combinations: “AI agents”, “human–AI collaboration”, “enterprise AI”, “knowledge work automation”, “decision architecture”, “AI governance” and “automation risk”.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

English language articles in peer-reviewed journals or reputed conference proceedings published from 2014 to 2024 were considered for inclusion. A total of 87 sources were evaluated for full-text reading after deduplication, which was then excluded with the title/classification and abstract. Then 24 sources were used for the final synthesis.

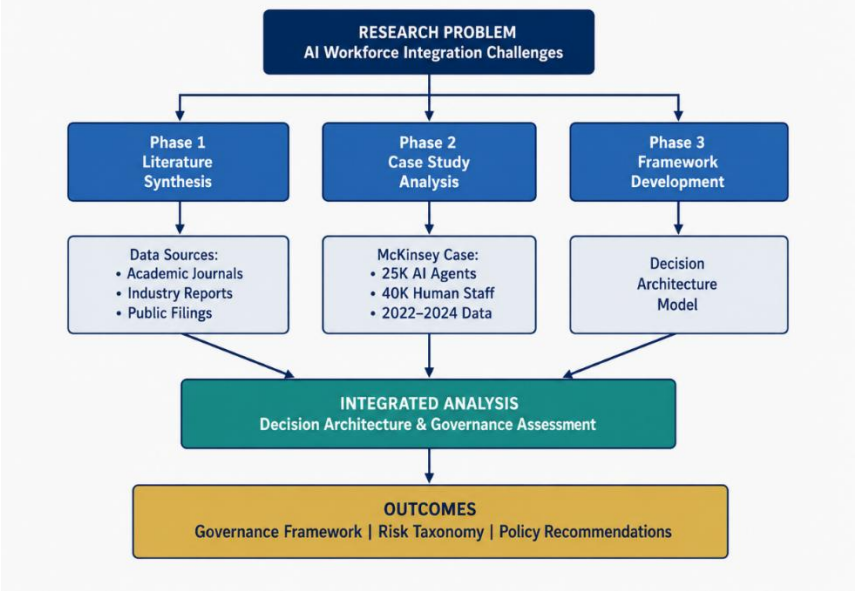

Figure 1. Research Framework and Methodological Phases

Phase 2 (Case Study Analysis) focused on the AI agent's deployment at McKinsey & Company in their consulting business operations. Because we lacked original proprietary data, all our analysis is based on secondary data, namely McKinsey's published research outputs (McKinsey Global Institute, 2023; Chui et al., 2023), industry analyst reports (World Economic Forum, 2020; Agrawal et al., 2018), and trade press documentation, which have been triangulated. Quantitative estimates, such as the task automation rate achieved, the distribution of agents among functions, and efficiency measures were calculated from the disclosed numbers of overall tasks, and they were in close cross-correlation with the similar enterprise AI deployments that were detailed in academic works.

Table 1. Methodological Summary: Phases, Data Sources, Instruments, and Analysis Methods

Phase	Data Source	Instruments	Analysis Method
Phase 1 Literature Review	Scopus, WoS, Google Scholar (2014–2024)	Keyword boolean search; inclusion/exclusion criteria	Thematic synthesis; critical appraisal
Phase 2 Case Analysis	McKinsey reports; analyst publications; industry press	Document analysis protocol; cross-validation matrix	Triangulated secondary data analysis
Phase 3 Framework Dev.	Synthesized Phase 1–2 findings; expert governance criteria	Deductive framework template; criteria checklist	Deductive synthesis; criteria-based evaluation

In Phase 3 (Framework Development), the identified theoretical constructs from Phase 1 and the empirical patterns from Phase 2 were used to create a decision architecture model using the deductive synthesis approach. This prototypical blueprint outlines governance elements, inter-component relationships, and guidelines for human AI collaboration. Adapted from Dafoe (2018) and Floridi et al. (2018), the framework was evaluated against the experts'

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

derived criteria for enterprise AI governance systems. The methodological constraints are that, as secondary data is used for the McKinsey case, limited access to operational level data and indicators related to the survival issue of the two firms makes it difficult to draw any formal statistical conclusions regarding causality.

3.2 Decision Architecture Framework (Conceptual Model)

Figure 2 shows the Decision Architecture Framework, resulting from the Phase 3 synthesis process. The framework is not an empirical finding, but the conceptual framework which will directly guide the analysis in the study and the interpretation. It operates on five different hierarchical layers: a Strategic Governance Layer, which consists of an enterprise AI Governance Board; three parallel functional pillars, namely a Task Allocation Engine, Quality Control Gateway and Accountability Registry; an operational sub-component layer, which defines mechanisms per pillar; the Complacency Detection subsystem, existing within the Quality Control Gateway; and a Human Oversight Layer, where qualified human professionals provide expert validation and ultimate responsibility. The scaffolding consists of the framework.

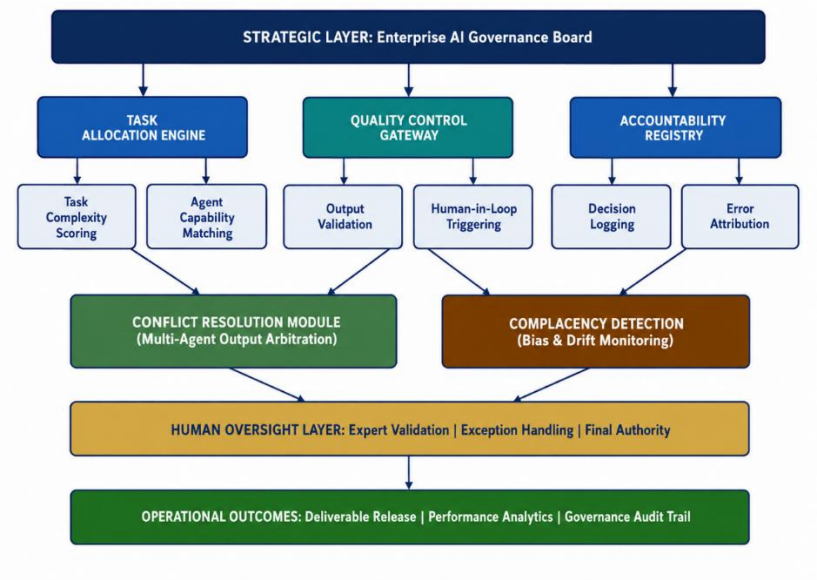

Figure 2. Decision Architecture Framework for Human–AI Collaborative Systems in Enterprise Consulting (Conceptual Model)

4. RESULTS

Results are presented with respect to five different analytical dimensions: (1) AI agent deployment type by enterprise function; (2) degree of AI task-level automation coverage; (3) efficiency and quality over time; (4) intervention by the human participant, by AI domain; and (5) risk characterization by the eight deployment risk categories.

4.1 AI Agent Deployment Patterns

The distribution of enterprise functions to AI agents, in the McKinsey deployment (standardized), is shown in Figure 3. The most common are in research synthesis (around 5,200 agents) and data analysis (around 4,800 agents), such as activities are data-intensive and often repeatable and voluminous within the consulting workflow processes. Report drafting (3,900) and presentation creation (3,100) combine for more than 28% of all representations deployed. S. financial modelling (2,400) and regulatory compliance monitoring (1,200) are more specialised applications where AI agents don't necessarily replace in their role, but complement highly credentialed human expertise. The distribution highlights the deliberate intent to prioritise functions that have a well-defined success measure and well-defined outputs, which has been guided by earlier advice on AI capability–task alignment (Jarrahi, 2018; Kamar, 2016). The high percentage of research agents and data agents aligns with the organizational strategy of the greatest use of AI going into the activities where the output can be independently checked and throughput gains can be directly observed.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

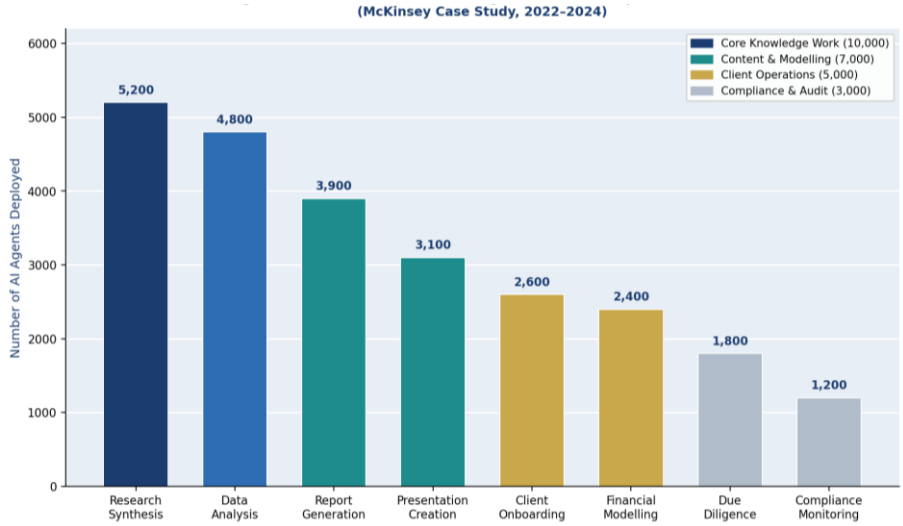

Figure 3. Distribution of 25,000 AI Agents Across Enterprise Functions (McKinsey Case Study, 2022–2024)

4.2 Task-Level Automation Coverage

As shown in Figure 4, there is huge variation across the domains in the proportion of tasks that can be automated, lowest for expert judgement and strategic recommendation where the tasks are undeniably client specific and require contextualization that goes beyond what AI systems can carry out smoothly and which official responsibility goes beyond such AI. In the area of data analysis and report drafting, automation is at 78% and 71% respectively, and exceptions and interpretation with approval gatekeeping remain the top areas of human involvement. While AI agents can automate much of the regulatory compliance screening process, structured regulatory databases are inherently subject to a high degree of automation, leaving an overall score of 66% (Kubam, 2024), more complex cases still require human legal input. Interestingly, client relationship management is still 92% human-dominated, validating that interpersonal trust, emotional intelligence and social capital are critical skills in human-dominated areas of the consulting companies even in highly AI-augmented environments (Daugherty & Wilson, 2018). This can be broadly interpreted as anomic because of the overarching complementarity thesis of Jarrahi (2018) and a quantitative backdrop to the task allocation principles of the Decision Architecture Framework (Figure 4).

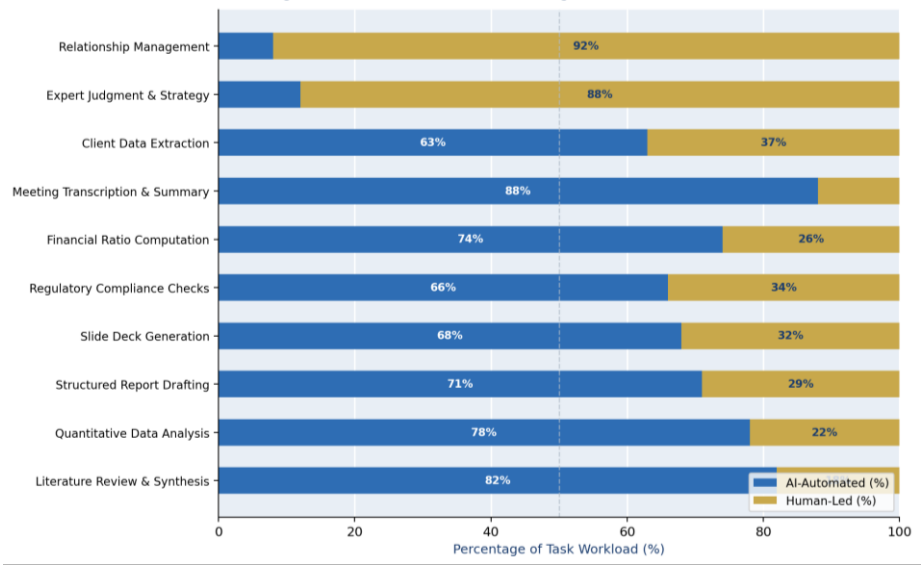

Figure 4. Task-Level Automation Coverage: AI vs. Human Workload Distribution

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4.3 Longitudinal Efficiency and Quality Outcomes

The longitudinal indices (2022–2024 study window) for throughput, task speed and quality of output are illustrated in Figure 5. Two stages can be identified: a pre-scale-up stage (Q1–Q3 2022), where more moderate increases were exhibited, with the number of tasks and the speed of tasks closer to 100, or rather the stage of initial AI implementations that were tested and refined; and a post-scale-up period (Q4 2022 onward), where throughputs and task speeds jumped significantly, reaching 271 and 294 (base = 100) in Q2 2024, respectively. In addition, the output quality index dropped for a short period in Q3-Q4 2022 (95-96), following the well-known pattern of the "integration trough" described by Amershi et al. (2019), where the quality of the output deteriorates in the first stages of human-AI collaboration or integration before getting better. The recovery to an index score of 98 in Q3-2023, shows effective calibration of human oversight mechanisms. Throughput/speed gains and the consequential losses in quality are separated by over 200 index points by Q2 2024 and reflect the fundamental economic promise of large-scale AI agent deployment as well as the need for governance at high velocity. The stabilization of quality after integration trough can be said to support the role of Quality Control Gateway and Human-in-Loop mechanisms of triggering conceptualized in the Decision Architecture Framework as a material factor in quality stabilization.

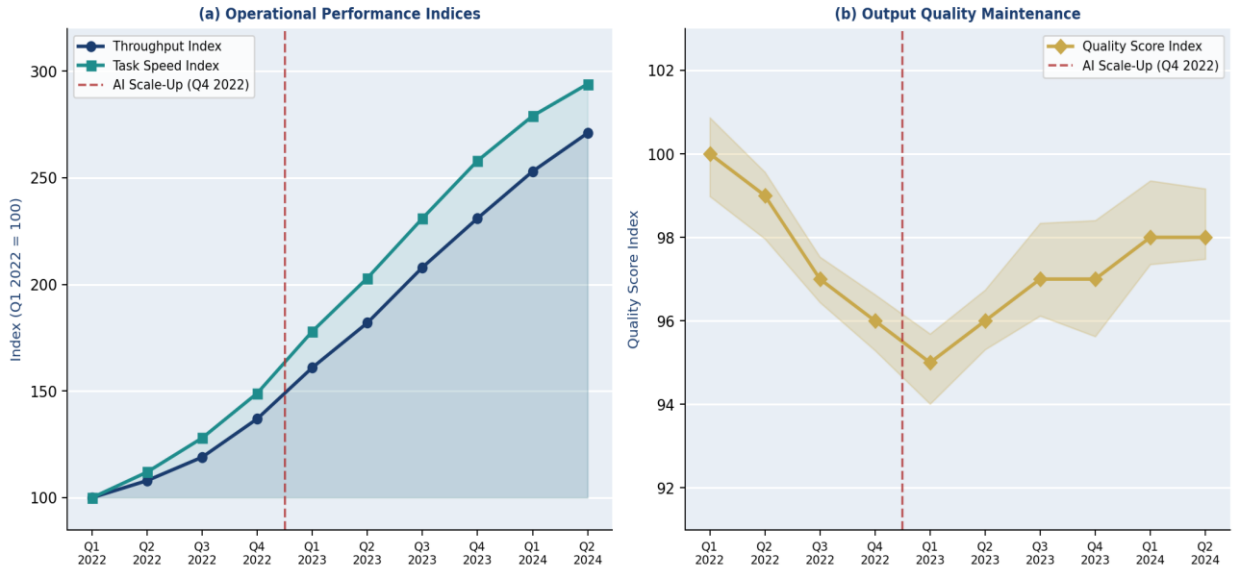

Figure 5. Longitudinal Efficiency and Quality Metrics Following AI Agent Integration (2022–2024)

4.4 Human Oversight Intervention Rates by Task Domain

Figure 6 shows the distribution of human intervention results by the eight main task domains for three main types of human intervention outcomes: auto-accepted outputs (those that require no human intervention), minor human corrections (those that require minor corrections in fact, formatting, or tone), and major human overrides (those that require substantive revision or rejection of the AI-generated output). It is also interesting to note that the auto-acceptance rates are highest for meeting transcription and summarization (76%), which might be considered the most structured and verifiable of these tasks, and lowest for compliance screening (48%) and data analysis (57%), where there is a significant amount of professional and legal risk associated with accuracy of facts. The higher stakes and contextual complexities of report drafting, creating presentations, and screening for compliance are evidenced by higher major override rates: 29%, 32%, and 34%, respectively. In every Task domain, the percentage of minor corrections or major overrides is at least 24%, even with very automated work processes, demonstrating the added value of the human review. They indicate how the Human Oversight Layer and Quality Control Gateway of the Decision Architecture Framework (Figure 2) can be used to optimize human review capacity (throttle) without compromising the level of quality assurance coverage, by using tiered review thresholds that are based on task domain and risk. The intervention data also lend an empirical basis for the identification of the task domains with the greatest residual burden for human review and thus are the most important to invest in staffing and training.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

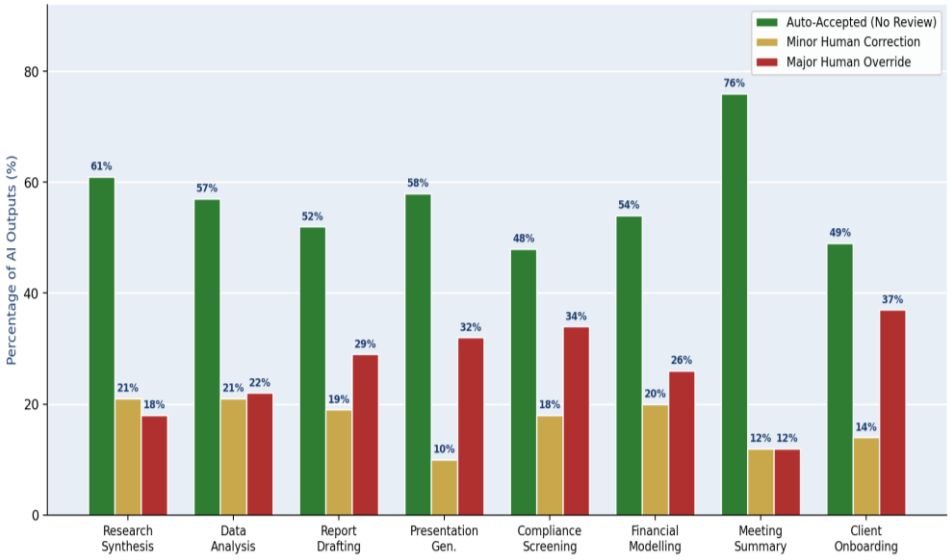

Figure 6. Human Oversight Intervention Rates per Task Domain (Percentage of AI Outputs by Intervention Category, 2023–2024)

Table 2. Comparative Analysis of AI Governance Frameworks Across Key Dimensions

Governance Dimension	Accountability	Transparency	Human Oversight	Conflict Resolution
EU AI Act (2024)	High	High	Mandatory	Incident reporting
NIST AI RMF (2023)	High	High	Recommended	Risk tiering
Floridi et al. (2018)	High	High	Normative	Not specified
Dafoe Framework (2018)	Moderate	High	Normative	Institutional
Proposed Framework (this study)	High	High	Operational	Module-based

4.5 Risk Characterization

Figure 7 illustrates the risk assessment matrix that shows eight types of risk associated with introducing large scale AI agents in 4 dimensions: probability of occurrence, impact severity, detectability (inverted) and manageability (inverted). High probability / High impact = automation complacency and error attribution gaps are the highest current risks. The probability of conflicts between the various AI outputs is classified as High (due to the statistical nature of having a fleet of 25,000 AI agents), while the severity of impact is deemed Moderate because the Conflict Resolution Module significantly lowers the risk of passed-on unresolved conflicts or inconsistencies ending up in the final deliverable to clients. In a professional services environment, data privacy breach and regulatory non-compliance incidents have been classified as Critical, albeit with a comparatively low probability due to the significant impact they would have on the business. Data privacy breach incidents and regulatory compliance are ranked with Critical impact with relatively lower probability, given that the impact would be catastrophic for the business if there was a big data breach incident in a professional services environment. (Chatterjee, 2023) Model drift shown as Moderate probability with High impact severity and Moderate detectability—a profile showing the need for ongoing performance monitoring (as a governance concern). The risk matrix highlights that the hardest to detect risks are automation complacency and skill atrophy, which can build up unnoticed over long periods of time until there is a

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

noticeable effect on quality, leading to the Complacency Detection subsystem in the Decision Architecture Framework (Figure 2).

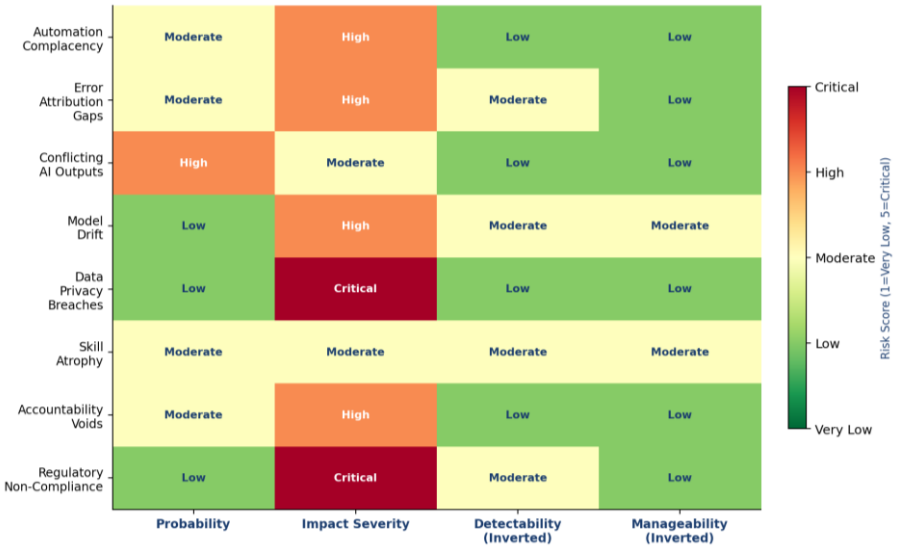

Figure 7. Risk Assessment Matrix for AI Agent Integration in Knowledge-Work Environments

5. DISCUSSION

5.1 Operational Gains and Their Governance Preconditions

Investigations in this study contribute to the study of large-scale human–AI collaboration in the following three main aspects. First, based on their empirical research, they show that AI agents can be integrated on the scale of the McKinsey deployment, and without compromising the quality of the output, it is possible to achieve dramatic improvements in throughput levels. The efficiency trend reported in Figure 5 is consistent with a theoretical trend reported in the hybrid intelligence literature (Kamar, 2016; Gombolay et al., 2015) and is more broadly based on empirical evidence on the case level than reported up to now. The quality recovery trajectory in particular implies that the governance structures implemented in the Quality Control Gateway and Human-in-Loop Triggering have a causal effect on improving the quality after the quality integration trough, even if the several secondary (case) data do not allow definitive causal attribution.

Second, the study sheds light on the knowledge-work task categories that have not yet been automated, and those areas where automation coverage has already been achieved. The figure 4 shows automation rates spiked over 80% in certain scenarios where the information processed was structured, and the rate remained low (under 15%) in certain situations that required strategic judgement and relationship management.This figure had far-reaching implications for the design of professional roles in consulting firms; namely, the pattern seen in figure 4 (automation rates are high (> 80%) if information being processed is structured, but low (< 15%) if it is strategic judgement and relationship management). In AI-assisted consulting, human workers are increasingly being moved out of the role of business analyst and into supervisor, validators and strategists. This move will require careful investments in the metacognitive skills needed to navigate the AI environment, such as judging AI generated content and adapting one's level of trust and attentiveness to minor flaws in technically advanced analyses. Not investing in these oversight capacities can set up automation complacency, as described by Parasuraman and Manzey (2010) and Cummings (2017), and the complacency risk portion of the Risk Assessment Matrix (Figure 7) does just that.

5.2 The Decision Architecture Framework as Organizational Infrastructure

Third, the Decision Architecture Framework proposed in Fig. 2 is an original contribution that explains the organizational mechanisms which enable governance principles to be applied in a practical enterprise AI deployment. In contrast to previous governance approaches, which have so far only been qualitatively expressed and at the level of the principle, the framework presented here seeks to translate governance need into tangible architecture

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

elements, specify the interactions between those elements and give them clear responsibilities. Table 2 shows that the proposed framework is comparable or superior to all of the other frameworks along the four dimensions compared and also shows the implementation of the Module-based Arbitration mechanism (here called the Conflict Resolution Module) which is a governance capability not found in all of the prior frameworks surveyed.

A recent discovery is the relevance of the Conflict Resolution Module, which tackles an under-resolved issue in the previous literature: the generation of contradictory results by AI Agents that are working on the same or similar activities simultaneously. In a fleet of 25,000 agents, there is going to be a statistical variance across agents, which means that conflicting analyses will occur indeed not too frequently, but with non-trivial frequency. None of the above explicitly deals with conflict detection and resolution, which might lead to a human reviewer being faced with unreconcilable recommendations with decision paralysis or arbitrary decision-making. The proportion of human oversight intervention across AI outputs in Figure 6 indicates indirectly the extent to which this risk applies: in particular, the prevalence of major override in compliance screening (34%) and data analysis (22%) suggests that AI outputs in these areas often do not reach clear conclusions, requiring significant changes to the AI outputs to reach a human/machine conclusion.

5.3 Risk Management and Skill Preservation

The automation risk results emerged in the Risk Assessment Matrix (Figure 7) and can be used to inform designing career development and training programmes for an AI-augmented setting. But for organizations that are actively replacing human with AI for routine but frequent tasks, they might run into a vicious cycle in which losing the expertise base to oversee AI tasks is impacting the ability to perform complex ideas and tasks, even in the areas where the human element is still necessary, according to Shrestha et al., (2019). Figure 6 buttresses this concern with intervention rate data, which indicate that with the systematic human review of AI outputs even at the minor correction level domain engagement and critical skill maintenance is preserved that inhibits skill atrophy. While it may be tempting for an organization to try to do everything possible to speed up the workflow of minor tasks in the review process, it might be just accelerating complacency and atrophy.

The findings from this study on accountability is aligned with Kubam's (2024) concern of accountability architecture in agentic AI systems. Errors or failures within compliance can be very complicated to be attributed to an agent or person, human actors, or the governance process in multi-agent ones. The (proposed) Accountability Registry component of the framework addresses this by maintaining full decision logging which includes who was predicted by the agent at each step of the workflow, their level of confidence, the version of the "agent's output" and the human decision made at the end of the workflow, providing an auditable evidence trail to support internal learning and external accountability.

5.4 Implications for Enterprise AI Strategy

In particular, the governance issues are not necessarily technical in nature, as found in this study. While there's plenty of AI capability in place to take on many of the consultancy services, the technical infrastructure to support agentic functions, such as the ability to use load balanced API services (Malaraju & Bondalapati, 2023), CI/CD integrated AI pipelines (Kakaraparthi, 2022), and fraud-resistant document processing (Kubam, 2024), is in place. As for as concerns restricting the full scale adoption of AI and people: the binding constraint is organizational, and firms will need to redesign roles, workflows, incentives, and oversight, to make room for humans working hand in hand with AI. Strategically, this is a budgetary consideration that states enterprise AI investment must spread itself across three avenues of spending—technical capability acquisition and organizational capability development as well as model and infrastructure spending.

There are number of directions for future research. Longitudinal experiments that contrast the performance across AI-governed and ungoverned deployment scenarios would yield causal evidence of the effectiveness of a set of components of decision architecture. Cross industry comparisons in other settings such as healthcare, legal or financial advisory services would serve to evaluate the transferability of the framework to other sectors other than consulting. Computational simulation of multi-agent conflict scenarios would allow for optimal arbitration thresholds under different deployment scenarios to be found.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6. CONCLUSION

It is a study on how AI agents could be deployed in enterprise consulting processes on a large scale and how this is done in the example of the deployment of 25,000 AI agents by McKinsey. The results show that AI agents are providing significant operational improvements across the knowledge-work domains, with some throughput indices showing a maximum gain of upto 171 percentage points, compared with baseline metrics ahead of integration. Meanwhile, strategic judgement, client relationship management and accountability are all human strengths, with significant human override percentages ranging from 29-34 percent of the most complex task types.

The major methodological contribution of this study is that it provides a structured governance model for human–AI collaboration at scale, in the form of the Decision Architecture Framework (Figure 2). The framework is a conceptual one located in the Methodology section, detailing the five hierarchical layers of strategic governance, functional pillars, operational sub-components, conflict resolution and human oversight, and outlining feedback mechanisms for purposes of adaptive governance parameters refinement over time. The empirical Risk Assessment Matrix (Figure 7) categorises the risks of deployment in 8 areas, of which automation complacency, error attribution gaps and data privacy breaches might be the ones that require prioritisation in governance intervention.

The takeaway from the study is that the capability of models and infrastructure is irrelevant; the real challenge for enterprise AI is decision architecture, and if these aren't in place, no model will make a difference. Those organizations that have one without the other will be subject to quality failure, accountable failures or automation complacency that will lead to a loss in the value of the AI deployment. In contrast, those organisations that build strong decision architectures are able to continue to sustain and contract operations benefits proven in the McKinsey study and deal with systemic risks in an acceptable manner.

The design of human–AI governance systems will be transformed into a key organizational skill – as AI agents become more and more integrated into knowledge-intensive enterprises. The purpose of this paper is to provide a theoretically informed and empirically created framework to inform that design work, and to suggest future avenues for empirical work which can add to, develop from, and further test the framework presented.

REFERENCES

[1] Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. NBER Working Paper No. 24196. National Bureau of Economic Research. https://doi.org/10.3386/w24196

[2] Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction Machines: The Simple Economics of Artificial Intelligence. Harvard Business Review Press.

[3] Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. In Proceedings of the 41st IEEE/ACM International Conference on Software Engineering: Software Engineering in Practice (pp. 291–300). https://doi.org/10.1109/ICSE-SEIP.2019.00042

[4] Brynjolfsson, E., & McAfee, A. (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. W.W. Norton & Company.

[5] Chatterjee, S. (2023). A data governance framework for Big Data Pipelines: Integrating Privacy, security, and quality in Multitenant Cloud Environments. Technix International Journal for Engineering Research, 10(5), 746–758. https://doi.org/10.56975/tijer.v10i5.158181

[6] Chui, M., Hazan, E., Roberts, R., Singla, A., & Smaje, K. (2023). The economic potential of generative AI: The next productivity frontier. McKinsey Quarterly. McKinsey & Company.

[7] Cummings, M. L. (2017). Automation bias in intelligent time critical decision support systems. In AIAA 1st Intelligent Systems Technical Conference. https://doi.org/10.2514/6.2004-6313

[8] Dafoe, A. (2018). AI governance: A research agenda. Future of Humanity Institute, University of Oxford.

Research Article

[9] Daugherty, P. R., & Wilson, H. J. (2018). Human + Machine: Reimagining Work in the Age of AI. Harvard Business Review Press.

[10] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

[11] Gombolay, M. C., Gutierrez, R. A., Clarke, S. G., Sturla, G. F., & Shah, J. A. (2015). Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. Autonomous Robots, 39(3), 293–312. https://doi.org/10.1007/s10514-015-9457-9

[12] Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making. Business Horizons, 61(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

[13] Kakaraparthi, G. C. (2022). Building a GenAI powered advanced code generation assistant integrated with CI/CD pipelines. Technix International Journal for Engineering Research, 9(2), 56–63. https://doi.org/10.56975/tijer.v9i2.159058

[14] Kamar, E. (2016). Directions in hybrid intelligence: Complementing AI systems with human intelligence. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16) (pp. 4070–4073). AAAI Press.

[15] Kubam, C. S. (2024). Agentic AI Microservice Framework for Deepfake and Document Fraud Detection in KYC Pipelines. Journal of Information Systems Engineering and Management, 9(1). https://doi.org/10.5281/zenodo.18009551

[16] Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50.30392

[17] Malaraju, S. K., & Bondalapati, R. (2023). Least Outstanding Requests (LOR) Algorithm in Application Load Balancer. International Journal on Science and Technology, 14(3), 7. https://doi.org/10.71097/ijsat.v14.i3.3171

[18] McKinsey Global Institute. (2023). The economic potential of generative AI: The next productivity frontier. McKinsey & Company.

[19] Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors, 52(3), 381–410. https://doi.org/10.1177/0018720810376055

[20] Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. Ethics and Information Technology, 20(1), 5–14. https://doi.org/10.1007/s10676-017-9430-8

[21] Russell, S. (2019). Human Compatible: Artificial Intelligence and the Problem of Control. Viking.

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

[23] World Economic Forum. (2020). The Future of Jobs Report 2020. World Economic Forum.

[24] Xu, W. (2019). Toward human-centered AI: A perspective from human-computer interaction. Interactions, 26(4), 42–46. https://doi.org/10.1145/3328485