

Making AI Work at Scale: Operating Model and Governance Design for Accountable AI-Augmented Decision-Making in Complex Organisations

By Bonnie Pui Shan So

Introduction

AI has progressed rapidly from an innovation concept to an operational reality. Organisations are already using it for forecasting, customer service deployment, content generation and re-use, knowledge retrieval, anomaly detection and a wide range of operational decisions. In many cases, AI increases speed and effectiveness. What is less clear is how organisations can leverage it at scale without introducing new risk, ambiguity or fragility.

When AI is added to the mix not only as an assistive tool but also as a means to reduce labour costs, this question becomes more complex. However, in practice it is a magnet for boards and executive teams, who see it as cheaper, quicker and more scalable than human labour. This creates an implicit or explicit expectation in some organisations that AI will replace parts of the workforce, reducing headcount, or restricting the scope for professional judgement. That financial logic is understandable, but it also adds complication to responsible implementation.

The organisational conversation shifts when AI is framed solely as cost-saving replacement for humans. Employees become more anxious and less trusting. Managers may exaggerate what the technology can do safely. Governance can be seen as an obstacle to speed and innovation rather than part of the design. At the same time, accountability becomes much more difficult to pin down because the organisation can still expect humans to bear the risk of decisions they no longer shape in any significant way.

These preliminary findings report arise from that practical tension. It is questioning why AI often excels in pilots but fails to become embedded in accountable, repeatable enterprise operations (particularly within large, regulated and operationally complex

organisations). The key argument is that this problem cannot be reduced to model performance alone. It is also a question of operating model design, governance maturity, decision rights, and workforce strategy.

In other words, the question is not just if AI works. It is whether organisations are rethinking work, authority and accountability in a manner that enables AI to scale as it should.

Context and Rationale

AI adoption is accelerating across industries, but enterprise-scale use remains uneven. Most organisations will be able to identify up-skilling pilots, local use cases or departmental wins. Fewer can demonstrate that AI has become integrated into normal operating practice in a way that is stable, auditable and defensible.

One reason is that AI enters organisations faster than the surrounding structures can adapt. Policies lag behind use. Decision rights remain vague. Escalation routes are unclear. Roles are only partly redesigned. That's why AI becomes one of several components of workflows and does not seamlessly integrate into an organisation's accountability architecture.

The second reason is more political and economic. In many organisations, AI is not used in a vacuum. It is doing so amid a backdrop of cost pressure, productivity expectations and high level interest in workforce reduction. Boards and C-suites like to see AI as an opportunity to do more with fewer people. That ambition is sometimes stated explicitly; other times it lurks behind the language of efficiency, transformation or simplification. Either way, it matters.

This has consequences for implementation. If the workforce sees AI primarily as a path to replacement, trust erodes. Employees may buck openly, acquiesce superficially, or use tools informally outside official governance. Asking staff to "own" AI-supported decisions (at the same time as taking away their discretion, time or power to question

outputs) This formulates a contradiction of social structure: civil liability remains on paper while human agency is reduced in practice.

That contradiction is most explicit in regulated environments. But if you work in sectors such as finance, insurance, healthcare, infrastructure or technology-enabled services, it may not be enough for the organisation to tell customers what decision was made — it must also explain how and why that decision was reached. AI can lead to better productivity, but it does not obviate the requirement for traceability, reviewability and defensible authority. If anything, the more forceful the drive toward automation, the more carefully designed those conditions must be made.

The pilot-to-scale gap, this study posits, is not primarily a technical one. It is a organisational design problem governed by governance, operating models and incentives. Whereas senior leaders see AI mostly as an incremental labour substitute, the risk intensifies further because the push to automate can outstrip the organisation's ability to govern responsibly.

Research Problem

The core problem this study addresses is the gap between successful AI pilots and responsible enterprise-scale use.

AI tools tend to work well in contrived circumstances with tightly overseen, highly sponsored and abnormally easy manual scrutiny. Yet when the same tools are scaled into normal operations by team, geography, and reporting relationship, the enabling conditions frequently evaporate. The organisation then faces a more fundamental question: not whether AI can create useful output, but whether the use of that output can be managed in a way that is repeatable and defensible.

Matters become murkier, particularly under an agenda of workforce-substitution using AI. Just because a pilot may appear to be successful - often through efforts or staffing in one area being decreased - does not tell you whether it is scalable. This can lead to the emergence of hidden costs later on such as oversight-related burdens, exception

handling and unhappy morale audit concerns and quality drift. Shorthand: What's cheaper at the pilot stage tends to turn more brittle during the enterprise one.

This study is interested not only in AI adoption, but in the organisational conditions under which AI-supported decisions can scale without hollowing out accountability. Existing literature helps explain AI capabilities, governance principles, and automation–augmentation tensions, but there remains a gap in practitioner-ready guidance on how to redesign roles, workflows, escalation paths, and oversight structures when AI becomes materially influential in decision-making.

Research Questions

This research is governed by 3 research questions:

1. **Why does AI often fail to progress beyond successful pilots in large, regulated organisations?**
2. **How do operating-model choices, including role redesign, workflow integration, decision rights, and escalation structures, enable or constrain responsible AI scale?**
3. **How do governance maturity, governance lag, shadow AI, and workforce substitution pressures influence the accountable use of AI across complex organisational settings?**

These questions are supported by provisional propositions:

- operating-model alignment is positively associated with responsible AI scale;
- governance maturity strengthens the link between technical capability and practical scale;
- governance lag and shadow AI are negatively associated with reliable and defensible scale;
- strong labour-substitution pressure may weaken trust, distort governance, and undermine sustainable scaling if not balanced by clear accountability design.

These propositions remain exploratory and will be refined through fieldwork.

Practical and Conceptual Framing

A key issue in this study is the distinction between augmentation and replacement.

Within much of the AI management literature, augmentation is framed as the more realistic and responsible course in complex environments. That is compelling in theory, especially where human judgement continues to matter. Organisational reality, however, is often messier. Senior executives are often under pressure to improve margins, reduce overhead, and deliver demonstrable returns. In those circumstances, it is easy to cast AI less as a partner in the workforce than as a substitute in that work. That framing is important because it determines behaviour.

Where AI is framed more as augmentation, organisations are more likely to spend on training, redesigned work, calibrated trust and human oversight. If it is framed predominantly as a substitute, organisations may underinvest in those things because the human component is viewed as a temporary load rather than an ongoing element of the system.

This tension also resonates with lessons from broader digital transformation. Managing Large Scale change is Difficult: Too often, large scale transformation efforts fail when leaders become too preoccupied with technology or a central plan at the expense of delivery realities, user needs and organisational resistance. The same pattern shows itself in AI. Purchasing tools is simpler than reengineering workflows. Announcing automation ambition is easier than building the governance routines to fix it. Counting the cost savings is easier than measuring whether decisions are auditable and defensible.

Well, this even relates to leadership thinking. In much of the literature, AI-fueled leadership is articulated in terms of accelerated decision-making and deeper insights. True, but leadership here is also taking responsibility for when not to automate, which

means human judgement must still play a meaningful role within, and for making sure the workforce does not become an expense line item to engineer out of existence. Responsible AI at scale in complex organisations is more than being efficient. It is also about legitimacy, trust and long term organisational capability.

Methodology and Ongoing Action

This study uses a mixed-methods design with a strong practitioner orientation.

Qualitative strand

The qualitative component employs purposive, theory-led elite sampling. Participants are invited not only for their technical AI expertise but also for their formal accountability role, governance authority, or oversight perspective. This encompasses leaders in risk, operations, digital and data governance, enterprise technology, audit, cybersecurity and executive management — along with regulatory or institutional perspectives.

This design encapsulates the idea that AI scale is not driven by a single channel within the organisation. A chief risk officer might prioritise accountability and challenge to models. Where a COO might be more hands on with workflow integration and operational redesign. Depending on the company, a CTO has to deal with scale architecture. Transformation narratives may miss weak spots, which internal audit can uncover. Insights from the boardroom or deputy CEO may show how cost pressure, workforce strategy and risk appetite drive AI decisions at the top.

The interview pool is purposely designed to capture these tensions across roles and sectors. As access is at least partly facilitated through professional networks, the study employed safeguards in the form of a structured protocol, thematic analysis and reflexive attention to researcher positionality.

Quantitative strand

This survey aims to discover trends in AI adoption, alignment to operating model, maturity of governance, shadow AI, workforce readiness and attitude towards decision

making aided by AI. It also includes things that draw attention to the disparity between formal governance and lived practice.

Another specific concern, especially relevant given the realities of this topic, is whether organisations publicly portray AI as augmentation while operationally seeking replacement. While this may not always be captured directly through specific survey wording, it is expected to bubble up indirectly through feedback on trust, planning for work redesign and the rise of shadow AI, management guidance and employee anxiety.

Action orientation

The goal is to create a pragmatic roadmap for organisations seeking to scale AI responsibly. The roadmap is provisional at this point, but initial reasoning points to a progression from:

1. supervised pilots,
2. fragmented and siloed use,
3. partial workflow integration,
4. accountable enterprise embedding,
5. adaptive operating-model maturity.

At each stage, the umbrella study is interested in structures of governance, but also whether there was leadership intent. An organisation that views AI primarily as a cost focused lever may progress through these stages quite differently from one that takes an approach of rethinking how humans and machines work together. This distinction may be crucial to whether scale is sustainable.

Challenges and Early Learning

Several challenges are already visible.

First, executive narratives on AI may be strategically polished. When strong internal pressure exists to reduce workforce cost, leaders might speak in the language of "innovation," "augmentation," and "responsible adoption." This isn't necessarily

dishonesty; it might reflect how hard it is to talk openly about labour substitution. It does mean that the research needs to listen carefully for tension between stated principles and operating realities.

Second, shadow AI isn't just a compliance issue. It may also indicate a mismatch between formal controls and real-world demand. Conversely, informal use may be motivated by job insecurity: the workforce might privately embrace AI to meet productivity demands in settings where governance and training are ambiguous.

Third and very importantly, accountability language can quickly become shallow. Many organisations call out that "humans stay accountable," yet the practical implications of this statement are highly diverse. But if it is assumed that a human will rubber stamp AI-shaped decisions with little time, information, authority or confidence to contest them, then formal accountability may not evidence genuine control.

Fourth, the workforce dimension cannot be simply an afterthought. If AI scale is linked with perceived replacement, then organisational trust becomes key. Loss of control leads to twisted adoption phenomena, low openness and implementation quality. That does not mean there is no place for automation. It means that AI governance must address the labour question honestly rather than couch it behind neutral language.

Expected Contribution to Practice

The intended contribution of this study is to pull together a practitioner's high-level playbook for accountable AI scale.

The playbook will help leaders respond to pragmatic questions like:

- What does "scale" mean when AI influences consequential decisions?
- Which decisions can be automated, and which must remain meaningfully human-owned?
- How should authority be distributed when AI recommendations become highly influential?

- What happens when the drive for labour efficiency conflicts with the need for review, challenge, and oversight?
- How should organisations interpret shadow AI and workforce resistance?
- What distinguishes an operating model that is scalable from one that is only efficient under pilot conditions?

The study is particularly pertinent for organisations under mounting pressure to innovate, cut costs and meet regulatory expectations all at the same time. Its potential value is in a step beyond abstract support for "responsible AI" and approaching a more pragmatic understanding of how governance, work design, leadership incentives and workforce strategy interact in practice.

Conclusion

AI solves real world problems at organisations already. But doing so at scale is harder than doing so in principle.

This work-in-progress highlights that the problem is not just a matter of technical capability. It is the challenge of developing operating models and governance frameworks that enable AI-backed decisions to be made in a repeatable, auditable and accountable way within everyday organisational practice. When AI is seen as a cheap replacement for human labour, rather than as an intricately regulated redesign of work, that challenge becomes even more pressing.

In large and regulated companies, responsibility at scale does not rely solely on model performance. You need to know how you want the system, or at least the process in your organisation — decision rights, workflow integration, escalation design, governance maturity and so on — to work together with human judgement and the credibility of leadership intent. The next stage of this research will enhance these concepts through interviews, survey data and the formulation of a practical roadmap to support organisations in their transition from AI promise to responsible operating reality.