Article

Not peer-reviewed version

Remote-Capable Knowledge Work Should Default to AI-Enabled Flexibility

Chaoyue He *, Xin Zhou, Di Wang, Hong Xu, Wei Liu, Chunyan Miao

Posted Date: 8 April 2026

doi: 10.20944/preprints202604.0561.v1

Keywords: remote work; flexible work; foundation models; generative AI; artifact capital; coordination costs; selective co-presence; future of work; algorithmic management; AI agents; knowledge work; organizational design

 doi:10.20944/preprints202604.0561.v1

Article

Remote-Capable Knowledge Work Should Default to AI-Enabled Flexibility

Chaoyue He [1,*], Xin Zhou [1], Hong Xu [1], Wei Liu [2] and Chunyan Miao [1]

1 Alibaba–NTU Global e-Sustainability CorpLab (ANGEL), Singapore
2 Alibaba Group, Hangzhou, China
* Correspondence: cyhe@ntu.edu.sg

Abstract

This position paper argues that remote-capable knowledge work should default to AI-enabled flexibility because the workflow-integrated foundation-model stack changes the coordination economics that once favored daily co-presence. By *foundation-model stack*, we mean systems that combine natural-language interaction, multimodal capture, long context, retrieval, transcription, translation, and increasingly bounded tool use inside everyday workflows. Their organizational significance is not generic automation but the accumulation of *artifact capital*: durable, queryable, reusable traces such as transcripts, summaries, decisions, tickets, code comments, and retrieval layers. The argument rests primarily on capabilities that are already widely deployed—transcription, summarization, retrieval, translation, drafting, and code assistance—with bounded agents treated as an amplifying but not necessary extension. Rather than eliminating the office, this shift supports *selective co-presence*, reserving in-person time for tasks with high tacitness, high coupling, or high relational stakes, including apprenticeship, conflict repair, trust formation, and early-stage synthesis. Because the same systems can also intensify surveillance, skill atrophy, and compute-related emissions, we outline a machine-learning research agenda centered on team-level evaluation, privacy-preserving memory layers, scaffolded AI for learning, carbon-aware routing, and pro-agency workflow design.

Keywords: remote work; flexible work; foundation models; generative AI; artifact capital; coordination costs; selective co-presence; future of work; algorithmic management; AI agents; knowledge work; organizational design

---

1. Introduction

Return-to-office debates are often framed as clashes of preference, but the deeper variable is coordination technology. As outlined in Figure 1, for remote-capable knowledge work, the central question is whether the workflow-integrated foundation-model stack now lowers the search, memory, handoff, and context-reconstruction costs that once made daily co-presence the safe organizational default. **Remote-capable knowledge work should default to AI-enabled flexibility, not by abolishing offices, but by making blanket attendance requirements increasingly hard to justify outside tasks with high tacitness, high coupling, or high relational stakes.**

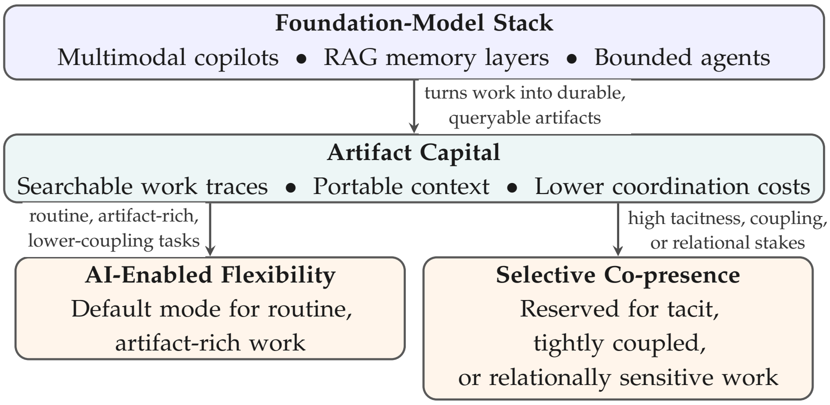


Figure 1. Overview: The foundation-model stack generates artifact capital that lowers the coordination costs of distance. This shifts routine tasks toward AI-enabled flexibility, reserving physical co-presence for highly tacit, coupled, or relational work.

This is a position paper: it synthesizes evidence from labor economics, organizational research, and deployed AI systems rather than presenting a new causal field experiment.

Offices became central not merely because managers valued visibility, but because co-presence historically reduced the transaction and coordination costs of finding expertise, aligning ambiguous tasks, transferring norms, and evaluating progress [1–5]. In this paper, *knowledge work* means work whose primary inputs and outputs are symbols, judgments, analysis, writing, coding, design, advising, and decisions rather than repeated physical transformation of materials [6,7]. We focus on *remote-capable knowledge work*: knowledge work whose core tasks can, for substantial periods, be executed through digital artifacts without continuous on-site physical presence [8]. The pandemic demonstrated that such work can be physically decoupled from the office, but it also revealed the frictions of distributed work: informal mentoring degrades, intra-organizational networks silo, and exploratory creative search can suffer under purely virtual communication [9–12]. At the same time, rigid attendance is increasingly costly in a workforce shaped by caregiving burdens, disability, and periodic disruption.

What changes the analysis is not "AI" in the abstract, but a specific technical configuration. By *foundation-model stack*, we mean workflow-integrated foundation-model systems that combine natural-language interaction, multimodal capture, long context, retrieval, transcription, translation, and increasingly bounded tool use [13–17]. Earlier digital tools typically improved one coordination stage at a time—search, storage, messaging, or automation. The foundation-model stack compresses production, comprehension, memory, and routing within the same workflow, which makes distributed coordination materially cheaper. Table 1 fixes the paper's terminology and scope.

 doi:10.20944/preprints202604.0561.v1

Table 1. Working definitions used throughout the paper. The position depends on *which* frictions AI reduces and *which* frictions remain.

Term	Working definition	Why it matters for the thesis
Knowledge work	Work whose primary inputs and outputs are symbols, judgments, analysis, writing, coding, design, advising, and decisions rather than repeated physical transformation of materials [6,7].	Sets the outer scope of the paper.
Remote-capable knowledge work	Knowledge work whose core tasks can, for substantial periods, be executed through digital artifacts without continuous on-site physical presence [8].	Narrows the claim to the part of knowledge work that is actually contestable.
AI-enabled flexibility	Meaningful choice over where and when work happens, combined with heavier asynchronous coordination and output-based evaluation, with the foundation-model stack reducing the search, memory, handoff, and context-reconstruction costs of distributed work.	This is the paper's proposed default, not an all-remote ideology.
Default operating model	A rebuttable presumption against blanket attendance mandates for remote-capable, artifact-rich tasks unless high tacitness, high coupling, high relational stakes, or evidence of harm justify in-person requirements.	Makes "default" operational rather than rhetorical.
Selective co-presence	Intentional in-person time reserved for onboarding, apprenticeship, trust formation, conflict repair, design sprints, and other high-tacitness or high-coupling tasks.	Predicts a narrower and more valuable role for the office, not its disappearance.
Tacitness	The share of task performance that depends on judgment, taste, context, or know-how that is hard to specify fully in advance.	Preserves a residual premium for in-person learning and judgment transfer.
Coupling	How strongly one person's progress depends on fast reciprocal adjustment with others rather than modular handoffs.	Preserves a premium for synchronization even when work is partly remote.
Artifact capital	Durable, searchable, reusable traces of work state — decisions, summaries, tickets, transcripts, code comments, rationales, exemplars, and retrieval layers — that let context travel across people and time.	Main mechanism through which AI lowers coordination cost.
Foundation-model stack	Workflow-integrated foundation-model systems combining natural-language interaction, multimodal capture, long context, retrieval, transcription, translation, and increasingly bounded tool use [13–17].	Distinct because it lowers the cost of producing, understanding, storing, and routing artifacts within the same workflow.

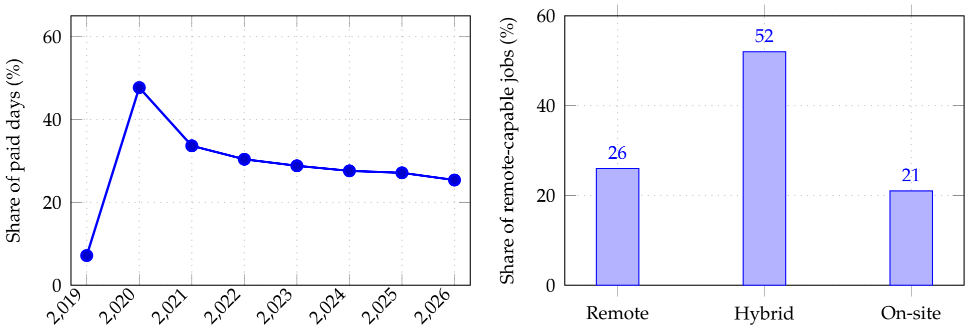
Here, "default" means a rebuttable presumption rather than a universal mandate: for remote-capable, artifact-rich tasks, blanket attendance should require task-specific justification or evidence that learning, mental health, reliability, or worker agency would otherwise be harmed. This is no longer a speculative argument from laboratory capability alone. By late 2024, 45% of Americans ages 18–64 had used generative AI, 27% of employed respondents used it for work in the previous week, and 10% used it every workday [18]. Follow-on 2025–2026 worker and firm surveys show large cross-country gaps and make clear that adoption depends not only on task exposure but also on management practices and whether firms actively encourage use [19,20]. The relevant question is therefore not whether these systems exist, but which coordination frictions they now make cheap enough to reorganize.

This paper advances four linked claims. First, flexible work has already stabilized as a structural equilibrium rather than a temporary pandemic artifact. Second, the foundation-model stack matters mainly because it creates and upgrades *artifact capital*, thereby reducing coordination costs rather than merely automating isolated tasks. Third, the deeper organizational effect is on latency, modularity, and control over the memory layer of the firm: these systems change how quickly work moves, how cleanly tasks decompose, and who can inspect or contest the resulting artifacts. Fourth, the implication is not uniform remote work but *selective co-presence*: offices remain valuable, but for narrower and more intentional purposes such as onboarding, apprenticeship, trust formation, conflict repair, design sprints, and tightly coupled creative work. The appendix provides falsifiable predictions, adoption notes, workflow cases, climate notes, guardrail design, a short related-work positioning section, and an extended literature map to make the position concrete rather than rhetorical. Appendix L situates the argument relative to adjacent work on agent harnesses, governed artifacts, and human–AI productivity measurement.

 doi:10.20944/preprints202604.0561.v1

2. Flexible Work Has Stabilized Rather Than Vanished

A useful baseline is not universal remote work but *selective flexibility* in remote-capable knowledge work. Many jobs cannot be done remotely at all, and even remote-capable jobs include tasks that benefit from periodic co-presence [8]. Still, three facts now make it difficult to treat attendance as the neutral default. Workers place real monetary value on flexibility [21]. Post-pandemic work patterns have stabilized above 2019 levels [22–26]. And hybrid is now the modal arrangement for many remote-capable jobs rather than a short-lived transition state [27,28]. Figure 2 summarizes that persistence.



(a) U.S. work-from-home rate

(b) Current U.S. location mix for remote-capable jobs

Figure 2. Flexible work has stabilized into a persistent hybrid equilibrium rather than reverting to 2019 levels. Panel (a) plots the U.S. work-from-home rate using the 2019 benchmark, annual averages of available series observations for 2020–2025, and the January–March 2026 average. Panel (b) details the current location mix for U.S. remote-capable jobs.

Persistence reflects both revealed firm behavior and worker demand. Workers place real monetary value on flexibility [21], and international survey evidence suggests that preferences for remote work remain widespread rather than uniquely American or pandemic [25,26]. The speed and shape of this equilibrium will vary across countries, sectors, and firm types because digital infrastructure, labor regulation, and management practice vary materially [19,20,26]. Once AI lowers the coordination penalty of distributed work, flexibility becomes part of the effective employment contract. Firms that insist on high-frequency attendance are then not merely choosing a culture; they are implicitly choosing a narrower talent pool, higher compensating differentials, or higher attrition risk.

The performance evidence is likewise more nuanced than either side of the public debate usually admits. Meta-analytic and review work already showed that telecommuting effects are heterogeneous, with benefits depending on autonomy, communication quality, and the fit between task design and work arrangement [29,30]. Since 2020, field studies have strengthened the case. In the Chinese call-center experiment, working from home increased measured performance and reduced attrition, although with career frictions for home workers [31]. Hybrid work at Trip.com improved satisfaction and reduced quits without harming performance [32]. Geographic flexibility raised output among patent examiners [33], and well-designed "smart working" increased both productivity and well-being [34]. Even studies that find costs typically point to mentoring, collaboration, or knowledge spillovers rather than to any generic inability to work at distance [9,35].

That evidence does *not* say that all jobs should become fully remote. It does say that rigid attendance is no longer the neutral baseline against which flexibility must plead for exceptions. Flexible work is already durable enough that the strategic question has changed from *whether* it is viable to *which frictions still prevent it from scaling*. AI enters at precisely that margin.

3. Why the Foundation-Model Stack Matters: Cheaper Coordination

The distinctive contribution of the foundation-model stack is not that it suddenly makes flexible work possible. Cloud software, search, version control, and messaging already did that. The discontinuity is that a workflow-integrated stack now lowers the cost of producing, understanding, storing,

 doi:10.20944/preprints202604.0561.v1

and routing artifacts inside the same routine. Natural-language interfaces lower self-service costs; multimodality turns speech, screenshots, slides, and mixed-format documents into machine-readable inputs; long context and retrieval make prior work more queryable; and bounded agents begin to move workflow state across tools [13–17]. By late 2024, work use of generative AI had already reached 27% of employed U.S. adults in the previous week, with estimated time savings equivalent to about 1.4% of total work hours [18]. Follow-on 2025–2026 surveys show large cross-country adoption gaps and highlight management encouragement as a meaningful complement to technical capability [19,20].

The recent wave has unfolded in layers rather than in one leap. The first layer was conversational assistance: workers could ask for a draft, explanation, rewrite, or code snippet in natural language, which made AI a practical general-purpose assistant rather than a specialist tool and sharply lowered the cost of self-service problem solving [18,36]. The second layer was workflow embedding and multimodality: systems increasingly appeared inside meetings, documents, productivity tools, and code workflows, while also parsing speech, screenshots, slides, tables, and mixed-format context [13,14,37]. The third layer was long context plus retrieval, which made decisions, tickets, repositories, and policy documents far more queryable across time [14,15]. The fourth layer is bounded agents: systems increasingly gather evidence, draft updates, and move workflow state across tools, so the gain is no longer only faster writing but lower follow-through latency [16,17]. Evidence on worker demand suggests that these coordination components are among the tasks workers most want AI agents to support [38]. Each layer removed a different reason people once needed the same room every day. Related systems work on harness engineering makes the same point from the agent side: tool mediation, runtime control, verification, and trace retention belong to the effective capability layer that organizations deploy, not merely to prompting style [39].

Crucially, the paper's core claim does not depend on frontier-level agent reliability. The strongest evidence today is for transcription, summarization, retrieval, translation, first-pass drafting, and code assistance [36,37,40,41]. Bounded agents amplify the argument by lowering follow-through latency, but the repricing of co-presence begins before agents become fully robust. What matters for AI-enabled flexibility is therefore not text generation in the abstract. It is the cumulative reduction of distributed-work frictions through artifact capital. Table 2 summarizes where the new stack substitutes for historical office functions and where co-presence retains an advantage.

We refer to the resulting stock as *artifact capital*: durable, searchable, reusable representations of work state such as summaries, decisions, design rationales, code comments, exemplars, translations, retrieval layers, and structured handoffs that reduce the cost of future coordination. The idea is consistent with classic accounts of the firm as a response to transaction and knowledge problems [1,2,5], with theories of tacit knowledge and knowledge creation [42–45], and with coordination research emphasizing expertise location, common ground, and shared memory [4,46–48].

A compact way to express the mechanism is given in Equation 1:

$$V(m) = B_{\text{trust}}(m) + B_{\text{learning}}(m) + B_{\text{creative}}(m) - \sum_{k \in K} C_k(m), \tag{1}$$

where $K = \{\text{search, handoff, memory, monitor, space}\}$. In this equation, m represents a specific work mode (e.g., asynchronous remote work, hybrid work, or dense co-presence), and $V(m)$ is the net organizational value of that mode. The B terms represent the benefits that historically favored the office: $B_{\text{trust}}(m)$ is the value of interpersonal trust formation, $B_{\text{learning}}(m)$ captures apprenticeship and tacit skill transfer, and $B_{\text{creative}}(m)$ denotes the output of tightly coupled creative search. The final term subtracts the sum of coordination costs, $C_k(m)$, across the set of organizational frictions K. AI mainly lowers these C_k costs for artifact-rich tasks by reducing the expense of finding prior decisions, reconstructing context, preparing handoffs, and producing inspectable outputs. It does much less to erase the B benefits of trust, apprenticeship, and tightly coupled search. The result is not "remote wins," but a repricing of where the office still earns a large premium.

 doi:10.20944/preprints202604.0561.v1

The deeper organizational effect is often on *latency* rather than on average task speed. Offices historically earned a premium because they collapsed waiting times: the missing answer, quick clarification, or informal escalation that prevented a task from stalling for a day. AI copilots and retrieval layers attack that tail latency directly by producing first-pass summaries, surfacing precedents, drafting follow-ups, and routing the next action before a colleague is free. Even modest task-level savings can therefore propagate into larger coordination gains when they shorten queues and reduce idle waiting.

Four channels matter most. First, **Context Portability**: AI makes context portable by turning ephemeral discussions into searchable summaries and action items, enabling enterprise retrieval over past specs, tickets, code reviews, and policies, and extracting structure from multimodal inputs like slides and forms—addressing coordination failures that stem less from missing talent than from missing retrieval and shared context [49–51]. Second, **Communication Compression**: AI compresses communication overhead by rewriting, translating, summarizing, and drafting, reducing both the blank-page and reconstruction taxes; evidence from the Microsoft 365 Copilot field experiment shows workers spending about three fewer hours (roughly 25%) less on email weekly without significant changes in meeting time, consistent with AI first lowering individually adjustable coordination burdens [37]. Third, **Expertise Diffusion**: AI diffuses expertise by delivering disproportionate gains to less experienced workers or those facing high setup costs, as seen in customer support (+15% productivity, especially for lower-skilled workers), software development (+26.08% completed tasks), and broader experiments where AI substantially narrows education-based productivity gaps—indicating that workers can access patterns and guidance without needing immediate proximity to experts [40,41,52,53]. Fourth, **Coordination Automation**: AI begins to automate coordination preparation and follow-through through agentic systems that collect evidence, prefill approvals, update systems, and draft plans, while humans retain control over judgment and exceptions; however, as a system technology, AI's value depends on complementary organizational redesign rather than standalone deployment [20,38,54–56].

The practical effect is cultural before it is fully macroeconomic. Conversational assistants lower the threshold for self-service problem solving; workflow-integrated multimodal copilots turn meetings and drafts into reusable artifacts; long-context retrieval reduces the gatekeeping value of sitting near the person who "knows where things are"; and bounded agents reduce follow-through costs across tools. The net result is a shift from "ask the person" to "query the work," from meetings as memory to meetings as escalation, and from commute rhythms to selective co-presence organized around high-coupling tasks. In the Microsoft 365 Copilot field experiment, treated workers reduced time spent on email and also spent less time working outside regular hours [37]. Survey evidence also suggests that the durable endpoint is more likely hybrid than permanently all-remote: workers continue to value flexibility, yet fully remote workers report worse well-being than hybrid workers [26,57,58].

Current scientific evidence also suggests that this wave changes execution before it fully changes occupations. In software development, AI already writes an estimated 29% of Python functions in the United States and is associated with a 3.6% increase in quarterly online code contributions, yet the measured gains accrue mainly to experienced, senior-level developers while early-career developers show no significant benefits [59]. That is exactly what one would expect if current systems are strongest at accelerating artifact production, review, and exploration rather than replacing judgment about architecture, integration, or exceptions.

Table 2. Where AI substitutes for office functions and where it does not; co-presence becomes more selective as its comparative advantage narrows.

Historical office advantage	AI-enabled artifact or capability	Residual case for co-presence
Fast clarification, status checks, and meeting follow-up	Meeting copilots extract summaries and action items; draft replies and context-aware retrieval reduce manual reconstruction of context	Ambiguous alignment, sensitive trade-offs, and conflict resolution
Institutional memory lived in conversations and in individual heads	Transcripts, semantic search, decision logs, and retrieval over prior work make more knowledge portable and reusable	Norm formation, contested decisions, and culture transmission
Expertise clustered around nearby senior workers	Copilots, exemplars, role-adapted explanations, and first-pass guidance lower the penalty for not sitting next to the expert	Apprenticeship, judgment calibration, and tacit know-how
Time-zone and language frictions slowed global teamwork	Translation, rewriting, and context compression make asynchronous handoffs cheaper	Real-time co-design for tightly coupled or novel work
Routine follow-through depended on nearby administrators or synchronous pings	Agents can prefill forms, update tickets, gather evidence, and route routine approvals across tools	Exceptions, accountability, and ambiguous policy trade-offs
Presence served as a proxy for availability and effort	Richer artifacts and machine-readable work traces make output-based management more feasible	Trust calibration — but also a major surveillance risk

4. Deeper Implications: Latency, Modularity, Control Rights, and Culture

Four structural implications follow from this shift toward artifact-based coordination. First, **Tail Risk Compression**: AI fundamentally alters the economics of tail risk in coordination by surfacing precedents, reconstructing context, and drafting follow-ups, thereby reducing reliance on physical co-presence and enabling compounding team-level productivity gains even when individual tasks save only minutes. Second, **Modularity Premium**: AI disproportionately rewards decomposable, legible work, meaning teams with clear ownership, stable interfaces, and retrieval-ready documentation benefit far more than those reliant on undocumented interactions—reinforcing that AI, as a complementary system technology, delivers the greatest value when workflows are redesigned around durable artifacts rather than merely adopting models [20,54–56]. Third, **Memory Layer as Strategic Asset**: as machine-readable artifacts become the substrate of coordination, governance shifts from physical arrangements (“who sits where”) to data control (“who can write, inspect, and retain artifacts”), enabling scalable flexibility when properly permissioned but risking software-enforced presenteeism if misused [60–63]. Fourth, **Cultural Encoding Shift**: organizational culture, once transmitted through informal in-person interactions, is increasingly encoded in what is summarized, cited, and made searchable, enabling calmer asynchronous work when designed well but potentially creating pressure for constant algorithmic visibility when mismanaged—reframing the debate from office versus remote culture to high-trust artifact cultures versus low-trust digital presenteeism [60–63].

Ultimately, the foundation-model stack represents a two-stage reduction in coordination friction. Generative copilots first lowered the cost of producing and comprehending artifacts. Now, bounded agents are lowering the cost of moving workflow state across tools. The cumulative effect is profound: a critical mass of knowledge work can now be governed by shared memory and supervised workflows, undermining the historical necessity of habitual, daily co-location.

5. A Task-Based Theory of Selective Co-Presence

The strongest version of this position is not “remote work wins.” It is that co-presence becomes a scarce input allocated where tacitness, coupling, and relational stakes are highest. A useful mental model is a 2×2 over *tacitness* and *coupling*, consistent with classic ideas about task equivocality and media fit [64], as depicted in Figure 3. AI helps most when the missing ingredient is explanation, retrieval, or first-pass structure, and least when the missing ingredient is trust, taste, conflict repair, or political judgment. A practical decision rule follows: remote-capable, artifact-rich tasks should default to AI-enabled flexibility unless tacitness, coupling, or relational stakes are high enough to justify selective co-presence.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0561.v1

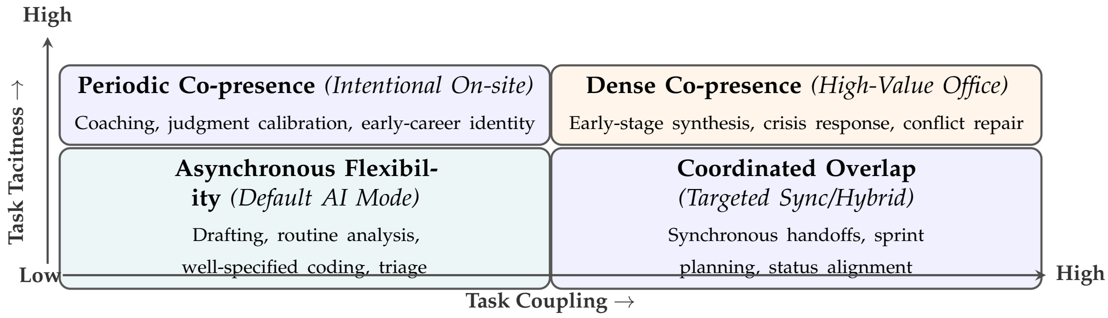


Figure 3. The Task-Based Matrix of Selective Co-presence. AI significantly lowers coordination frictions in the bottom-left quadrant. The physical office is retained as a targeted resource for tasks demanding high tacitness, high coupling, or both.

Low tacitness, low coupling. These tasks are the clearest candidates for asynchronous flexibility: drafting, routine analysis, well-specified coding, triage, document revision, and common customer interactions. They are already artifact-mediated, and AI further lowers their search, handoff, and setup costs. **Low tacitness, high coupling.** These tasks still benefit from synchronization, but not necessarily from daily co-presence. Hybrid blocks, coordinated overlap windows, and AI-supported handoffs are often sufficient. **High tacitness, low coupling.** These tasks need judgment transfer more than constant interaction: coaching, calibration, some forms of review, and identity-forming early-career work. The right response is usually periodic and intentional co-presence rather than blanket attendance. **High tacitness, high coupling.** These are the strongest residual cases for dense co-presence: early-stage team formation, crisis response, sensitive negotiation, conflict repair, onboarding waves, and creative synthesis under ambiguous objectives.

This framework explains why negative findings about remote work do not refute the thesis. If remote work reduces spontaneous cross-team ties or creativity [9,10,12], the implication is not that work should be colocated. It is that firms should purchase co-presence when they need bridging ties, joint search, or repair. Likewise, if proximity helps training and long-run skill formation [11], firms should design richer in-person windows for novices rather than impose uniform daily attendance on everyone. Flexible work becomes durable only if evaluation systems stop penalizing it through visibility bias or flexibility stigma [65–67]. That is the sense in which "default" is used throughout the paper: a rebuttable presumption, not a mandate.

6. Alternative Views

The claim that remote-capable knowledge work should default to AI-enabled flexibility is an empirically falsifiable position, not an ideological absolute. While the current trajectory favors this transition, the equilibrium could fail under specific socio-technical conditions (detailed fully in Appendix C). Below, we address the most significant counterarguments to this framework.

View 1: The Frontier Task Objection.

Some argue that AI will strengthen the rationale for the office because the hardest, most valuable frontier tasks remain highly interactive and ambiguous. If AI automates the routine execution, what is left is high-stakes, tightly coupled human brainstorming, which benefits immensely from physical proximity and high-bandwidth, in-room collaboration.

Response: This is a serious objection and likely true for specialized, frontier-research teams. However, it is insufficient as a general equilibrium for the broader knowledge economy. For most roles, the average day does not consist entirely of frontier brainstorming; it consists of triage, drafting, search, status updates, code review, and asynchronous handoffs. Those are precisely the tasks whose coordination tax is falling most quickly. The paper does not claim the office vanishes for frontier work; it claims the office is no longer the default for the massive volume of routine production.

View 2: The Digital Taylorism Objection.

A highly critical view posits that firms will use AI not to enable geographic flexibility, but to intensify monitoring and algorithmic management. If work becomes perfectly legible to software via "artifact capital," managers may use these systems to recreate the office in software through continuous surveillance, keystroke tracking, and automated performance ranking, negating any gains in worker autonomy.

Response: This is less a rebuttal of the paper's mechanism and more a competing—and highly concerning—implementation path. The capability to perfectly summarize and retrieve work state can absolutely be co-opted for surveillance. Our claim is therefore conditional: AI-enabled flexibility scales sustainably *only* when organizational governance tilts toward memory, support, and output-based accountability rather than toward continuous behavioral ranking. This makes privacy-preserving AI design a core labor issue.

View 3: The Skill Atrophy and Apprenticeship Objection.

Recent empirical evidence suggests that while AI drives short-run output gains, it can coexist with worse long-term learning. By outsourcing search, debugging, and drafting to AI, junior workers bypass the struggles through which deep expertise is built. Therefore, placing junior workers in remote environments with AI copilots risks generating a "lost generation" of talent that lacks foundational tacit knowledge.

Response: This concern is scientifically robust [68]. However, the solution is not a mandatory five-day return to the office for all employees. Instead, firms must distinguish execution support from apprenticeship. The framework of *selective co-presence* explicitly reserves intentional, in-person time for mentoring and judgment calibration. Furthermore, the ML community must build "scaffolded" AI that prioritizes explanation, Socratic dialogue, and calibrated uncertainty for novices, rather than just output generation.

View 4: The Mental Health and Social Isolation Objection.

A growing body of psychological and sociological research argues that remote work strips away the social fabric of the workplace, leading to increased rates of depression, loneliness, and anxiety. The office serves a latent function as a primary source of adult social connection; removing it degrades long-term mental health, well-being, and ultimately, organizational loyalty.

Response: This is a crucial limitation of the "all-remote" ideology, which is exactly why this paper argues for a *hybrid* flexible default rather than total virtuality. Fully remote workers do report higher rates of loneliness than their hybrid counterparts [57]. But forcing daily commutes for quiet, headphone-bound work does not solve loneliness; it repurposes time poorly. Selective co-presence is better aimed at social bonding, trust formation, and team building than either strict isolation or forced, performative attendance.

View 5: The Bounded-Agent Brittleness Objection.

Critics point out that AI agents hallucinate, lose context, and fail in enterprise environments. Because bounded agents cannot reliably move workflow state without errors, the coordination tax is not actually falling; rather, humans must remain tightly coupled and available to catch cascading AI mistakes and manage exceptions.

Response: This is true of early bounded-agent systems. But AI does not need perfect reliability to shift the equilibrium; it only needs to reduce follow-through costs relative to manual coordination, with humans reviewing exceptions and supervising the resulting artifacts.

View 6: The Macroeconomic Lag Objection.

Linked survey and administrative evidence frequently finds that rapid technology adoption coexists with null effects on aggregate earnings and recorded hours for years after deployment. Therefore, assuming that AI will immediately disrupt the deeply entrenched, macroeconomic real-

estate and labor structures of the office is premature.
Response: This caution disciplines the timeline, not the mechanism. Communication norms, meeting loads, and hybrid schedules can shift faster than aggregate wages or real-estate totals.

7. A Socio-Technical Research Agenda

If the argument advanced in this paper holds, AI-enabled flexibility is a concrete systems-and-evaluation problem for the machine learning community rather than a peripheral enterprise use case. The field needs benchmarks for handoff fidelity, context reconstruction, memory provenance, escalation quality, rollback after bad bounded-agent actions, after-hours spillover, and carbon per resolved task; Appendix Tables A5 and A10 summarize the capability classes and evaluation targets. We therefore propose the following socio-technical agenda so the shift enhances autonomy, learning, and sustainability rather than hardening into software-mediated presenteeism.

1. Team-Level and Longitudinal Evaluation Paradigms.

Current ML benchmarks overwhelmingly evaluate static, single-agent, or single-turn task completion. Flexible work, however, succeeds or fails through handoffs, shared memory, escalation, and coordination across time. The community should develop dynamic benchmarks that measure context reconstruction time, handoff fidelity, asynchronous onboarding speed, decision-provenance coverage, and retrieval accuracy over organizational history. We also need longitudinal field studies to track interruption load, meeting displacement, and whether AI adoption reduces coordination burden or merely displaces it into after-hours spillovers [37,57]. Related work on human–AI productivity measurement argues that deployment-relevant gains should be reported as time-to-acceptance under explicit acceptance tests, which is closely aligned with our emphasis on verification, rework, and governed handoffs rather than raw draft speed [69].

2. Scaffolded AI for Apprenticeship and Skill Formation.

AI-enabled flexibility changes who can access labor markets, potentially widening geographic matching and disability inclusion [33,70]. Yet the gains are distributionally uneven. If AI primarily accelerates experienced workers while automating the intermediate steps necessary for junior workers to learn [59,68], distributed work risks exacerbating a "novice penalty" and driving long-term skill atrophy. A critical research direction is designing *scaffolded copilots*—systems that prioritize calibrated explanations, Socratic dialogue, and reliable escalation to human experts rather than only final-output generation.

3. Privacy-Preserving and Interoperable Memory Layers.

If artifact capital becomes the substrate of coordination, the strategic asset is the organizational memory layer. The governance risk is that this layer devolves into a proprietary, panoptic surveillance apparatus [62,63]. We should build enterprise RAG and memory systems where provenance, access control, reversible summaries, and retention limits are formalized design constraints rather than compliance afterthoughts. Open schemas for interoperable artifacts are also essential to prevent cognitive and infrastructural lock-in within singular vendor ecosystems.

4. Carbon-Aware Routing and Systems-Level ESG Accounting.

The environmental case for flexible work is doubly material: it affects corporate real estate resilience and societal Scope 3 emissions [71–74]. However, the climate dividend is easily wiped out if commuting emissions are simply replaced by highly compute-intensive generative models and inefficient residential conditioning [75,76]. The Green AI community should prioritize carbon-aware model routing—sending routine summarization or retrieval tasks to smaller, lower-carbon models or off-peak compute regions [77]. Systems-level research should pair ML adoption with occupancy-responsive building tools so organizations consolidate physical footprints.

5. Pro-Agency Guardrails for Bounded-Agent Workflows.

As AI shifts from conversational interfaces to bounded agents that autonomously move workflow state across tools [16,17], the risk of unreviewable algorithmic management intensifies. Anti-surveillance and pro-agency design should therefore be treated as primary evaluation targets. The field should develop standards for inspectable agent actions, previewable multi-step plans, rollback mechanisms, and meaningful human override [78]. Evaluating these systems must answer not only whether they improve throughput, but whether they preserve worker discretion and trust.

8. Conclusion

The foundation-model stack does not make the office obsolete, but it does make blanket attendance a weak default for remote-capable knowledge work. By converting ephemeral interactions into queryable artifact capital, it reduces the historic frictions of distributed coordination and makes a large share of remote-capable knowledge work more asynchronous and portable across time and location. The resulting equilibrium is selective co-presence: intentional in-person time for apprenticeship, conflict repair, trust formation, and other high-tacitness or high-coupling tasks, paired with AI-enabled flexibility for routine execution. The central socio-technical challenge is to ensure that this transition strengthens worker autonomy, learning, and sustainability rather than hardening into software-mediated presenteeism.

Acknowledgments: This research is supported by the RIE2025 Industry Alignment Fund (Award I2301E0026) and the Alibaba–NTU Global e-Sustainability CorpLab.

This appendix is organized as follows. Appendix A narrows the claim; Appendix B states testable predictions; Appendix C specifies what would change our minds; Appendix D with Table A1 summarizes adoption evidence; Appendix E with Table A2 gives workflow cases; Appendix F with Table A3 lists failure modes and guardrails; Appendix G documents Figure 2; Appendix H with Table A4 summarizes climate evidence; Appendix I with Table A5 defines the capability stack; Appendix J with Table A6 stages the foundation-model stack over time; Appendix K with Table A7 lays out medium-run equilibria; Appendix L gives a short related-work positioning note; Appendix M with Table A8 maps the broader literature; and Table A10 closes with a compact evaluation checklist.

Appendix A. Clarifying the Claim

This paper does *not* claim five things.

1. It does not claim that every knowledge work job should become fully remote.
2. It does not claim that AI by itself guarantees productivity gains; organizational complements remain necessary.
3. It does not claim that co-presence loses value for onboarding, trust formation, conflict repair, creative synthesis, or highly coupled work.
4. It does not claim that flexible work is automatically equitable; unequal adoption and excessive surveillance could easily reverse its social benefits.
5. It does not claim that the office disappears. The claim is that the office becomes more selective, more episodic, and more defensible when tied to specific tasks rather than to habit.

The position is narrower and stronger: AI lowers the coordination costs that historically made rigid daily co-presence a broadly efficient default. That change shifts the burden of proof toward task-specific uses of the office.

Appendix B. Falsifiable Predictions and Empirical Agenda

The argument in the main paper yields at least six testable predictions.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0561.v1

1. **A declining co-location premium for artifact-rich tasks.** In occupations and teams where work already flows through documents, code, tickets, or customer records, the effect of co-location on measured output should fall as AI adoption rises.
2. **A shift from synchronous reconstruction to asynchronous handoffs.** AI adoption should reduce time spent rebuilding context, increase the use of summaries and decision logs, and raise the share of work carried through persistent artifacts even if meeting volume changes more slowly.
3. **Larger short-run gains for novices than for experts on routine coordination.** AI should compress some performance gaps on drafting, search, and first-pass problem solving while leaving open whether long-run skill formation still requires more proximity.
4. **Concentrated rather than uniform office use.** If the thesis is right, firms with high AI adoption should not simply choose "more remote" in a blanket way. They should reserve co-presence for onboarding, design sprints, conflict repair, and tightly coupled creative work.
5. **Heterogeneous gains by organizational maturity.** Teams with strong documentation norms, modular task structures, and clear ownership should realize larger flexibility gains than teams with weak process discipline.
6. **Divergence between autonomy-enhancing and surveillance-enhancing deployments.** Organizations that use AI primarily for retrieval, summarization, and support should see stronger retention and flexibility gains than organizations that deploy AI primarily for monitoring and ranking.

These predictions suggest a research agenda for ML, HCI, labor economics, organizational science, and sustainability measurement. The key outcome variables are not only task speed and accuracy, but also handoff quality, context reconstruction time, onboarding speed, career progression, after-hours workload, meeting displacement, carbon and resource intensity, privacy loss, and contestability of automated evaluation.

Appendix C. Conditions for Falsifiability: What Would Change Our Minds

The position advanced in this paper is not an ideological commitment to remote work; it is an empirical claim about coordination costs and organizational economics. Therefore, it is strictly falsifiable. We would revise our thesis that remote-capable knowledge work should default to AI-enabled flexibility if rigorous, longitudinal evidence demonstrates any of the following four conditions:

1. Systematic Reversion in High-Adoption Teams.

We would change our minds if empirical studies show that organizations with mature, fully integrated AI stacks (multimodal copilots, RAG memory layers, and workflow agents) systematically revert to requiring daily co-presence for routine, artifact-rich tasks. If the coordination tax remains high despite optimal technological deployment—meaning teams still cannot efficiently search, align, or hand off work without physical proximity—then our core mechanism mapping artifact capital to reduced coordination latency is fundamentally flawed.

2. Skill Atrophy Dominating Flexibility Gains.

This thesis relies on the ability to selectively use co-presence for high-tacitness apprenticeship while decoupling routine execution from the office. If long-term studies reveal that AI-assisted junior knowledge workers suffer catastrophic skill stagnation that cannot be mitigated by scaffolded AI or periodic, intentional co-location, the model fails. Specifically, if the loss in human capital and long-run innovation outweighs the immediate organizational gains in retention, global matching, and reduced commute friction, a mandatory return to continuous co-location for mentoring would be justified [68].

3. Unsustainable Mental Health Detriments.

While fully remote work is known to increase isolation, we posit that hybrid flexibility allows workers to reallocate energy and time from commutes into high-quality community or social connec-

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0561.v1

tion. If public health and sociological data conclusively prove that the baseline isolation of hybrid work directly causes a net deterioration in mental health that outweighs the burnout and stress prevented by eliminating forced daily commutes, the social sustainability argument collapses [57].

4. The Inevitability of Digital Taylorism.

We argue that artifact capital can and should serve as a supportive memory layer that enhances human autonomy. However, if evidence shows that the dominant, inescapable use of these systems across the market is to enforce ubiquitous algorithmic surveillance, keystroke tracking, and punitive micromanagement, we would reject AI-enabled flexibility as a positive organizational default. If the choice binary strictly reduces to physical autonomy in an office versus panoptic software surveillance at home, the office remains the safer harbor for worker agency [62,63].

Appendix D. Adoption and Diffusion Notes for the Foundation-Model Stack

As Table A1 summarizes, a useful synthesis is that adoption of the foundation-model stack appears fast enough to matter, heterogeneous enough to change inequality dynamics, and contingent enough to reward better management and process design. Those three properties together make it especially relevant to flexible work.

 doi:10.20944/preprints202604.0561.v1

Table A1. Selected evidence on adoption, diffusion, and heterogeneous returns in the foundation-model stack.

Source	Quantitative or substantive finding	Why it matters for this paper
Bick et al. [18]	By late 2024, 45% of Americans ages 18–64 had used generative AI; 27% of employed respondents used it for work in the previous week; 10% used it every workday; estimated time savings were about 1.4% of total work hours.	Shows that workplace use is already broad enough to matter for everyday coordination rather than only for frontier technical teams.
Bick et al. [19]	Worker and firm surveys from 2025–2026 document large U.S.-Europe adoption gaps and tie diffusion not only to workforce composition but to management practices and whether firms actively encourage AI use.	Supports the claim that flexible-work gains depend on organizational complements, not only on access to a model.
Cruces et al. [52]	In a randomized business-problem task, AI access reduced an education-based productivity gap from 0.548 to 0.139 standard deviations, closing about three quarters of the baseline difference.	Suggests that AI can level some execution barriers that previously made expertise location and proximity more important.
Daniotti et al. [59]	AI writes an estimated 29% of Python functions in the U.S.; quarterly online code contributions rise by 3.6%; measured benefits accrue mainly to experienced developers.	Indicates that AI already changes artifact production at meaningful scale while producing heterogeneous returns by experience.
Humlum and Vestergaard [79]	Rapid chatbot adoption in Denmark coexists with null effects on earnings and recorded hours larger than 2% two years after adoption, despite occupational switching and task restructuring.	Disciplines macro overclaiming: coordination equilibria and job design may shift before short-run wage or hours data visibly move.

Appendix E. Illustrative workflow Cases

As detailed in Table A2, the common pattern is that AI most reliably helps when work already leaves artifacts behind or can be made to do so. The residual demand for co-presence concentrates in the phases of work where tacit judgment, trust, and conflict management dominate.

 doi:10.20944/preprints202604.0561.v1

Table A2. Illustrative cases for how AI changes the feasible boundary of flexible work.

Function	Pattern under AI-enabled flexibility	Residual value of co-presence
Software engineering	Retrieval over prior pull requests, code search, test generation, ticket summarization, and bounded agents that prepare patches or update issue trackers reduce the setup cost of distributed development [41,59].	Architecture resets, codebase-wide conventions, mentoring of juniors, and conflict resolution around ownership or quality standards.
Customer support	Copilots surface precedents, summarize cases, draft replies, and standardize follow-up, making asynchronous handling more reliable [40].	Escalations, emotionally sensitive cases, and norm-setting around exceptions or difficult customers.
Writing, analysis, and communications	AI lowers the blank-page tax, rewrites for audience, translates across languages, and reconstructs prior context from long threads or document collections [36,37].	Message framing under high political stakes, sensitive stakeholder alignment, and training of junior analysts.
Professional services and back-office operations	Agents can gather evidence, prefill forms, route approvals, summarize policies, and update CRM or ERP systems across tools [17,38].	Accountability for exceptions, policy trade-offs, and relationship management with clients or regulators.
Product and strategy work	Meeting capture, decision logs, multimodal synthesis, and fast retrieval reduce the lossiness of asynchronous work.	Early-stage ideation, trust formation across functions, hard trade-offs under ambiguity, and coalition building.

Appendix F. Failure Modes, Rebound Effects, and Guardrails

Table A3 outlines these guardrails, which matter because the same technical capability can support either autonomy or control depending on how artifacts are governed.

 doi:10.20944/preprints202604.0561.v1

Table A3. Main failure modes for AI-enabled flexibility and practical guardrails.

Failure mode	Why it happens	Guardrail
Surveillance drift	Organizations use artifacts primarily for ranking and micro-monitoring rather than for coordination support.	Retention limits, purpose limitation, appeal rights, access controls, and a default presumption against productivity scoring from raw interaction traces.
Skill atrophy	Workers delegate exactly the steps through which they would otherwise learn search, reading, debugging, or synthesis.	Scaffolded use, explanation-first copilots, explicit escalation, protected practice tasks, and intentional apprenticeship windows.
Stale or misleading memory	Retrieval surfaces outdated, partial, or private material with confident fluency.	Provenance display, freshness checks, deletion and correction rights, and auditable citations to source artifacts.
Carbon rebound	Lower attendance does not translate into less office energy use; large-model defaults raise compute emissions.	Floor consolidation, occupancy-responsive HVAC, commute and telework accounting, and carbon-aware model routing.
Vendor lock-in of the memory layer	Valuable summaries, logs, and workflow state become trapped inside one product stack.	Exportable artifacts, interoperable APIs, model-agnostic storage, and procurement standards that require portability.
After-hours spillover	Faster asynchronous work silently shifts effort into evenings and weekends.	Track after-hours load, delay non-urgent notifications, and evaluate systems partly on whether they reduce rather than displace coordination burden.

Appendix G. Data Note for Figure 2

Figure 2 is included to make the paper less rhetorical and more concrete.

Panel (a) uses the Federal Reserve Bank of St. Louis FRED series `WFHCOVIDMATQUESTION`, which reports the share of full paid days worked from home in the United States. The figure uses the exact series benchmark for 2019 (7.1500), the mean of available 2020 monthly observations (47.6850), annual averages for 2021–2025 (33.6375, 30.3875, 28.8142, 27.5925, and 27.1167), and the January–March 2026 average (25.3767) [27]. The purpose is descriptive rather than causal: to show persistence far above the 2019 baseline even after the sharp retreat from the 2020 emergency peak.

Panel (b) uses Gallup's published distribution of work location among U.S. full-time employees with remote-capable jobs: 26% exclusively remote, 52% hybrid, and 21% on-site [28]. The published percentages sum to 99 because Gallup rounds category shares to whole numbers. This panel matters because it captures the more realistic equilibrium for remote-capable knowledge work. The live debate is increasingly not between "everyone remote" and "everyone back," but between hybrid/flexible defaults and renewed mandatory co-location.

Appendix H. Sustainability, ESG, and Climate Note

Table A4 summarizes selected quantitative findings that discipline the climate/ESG argument. The environmental case for flexible work is complex: while commuting emissions drop, home energy and compute footprints can offset these gains if not managed properly.

Table A4. Selected quantitative findings that discipline the climate/ESG argument.

Source	Quantitative result	Interpretation for this paper
Tao et al. [80]	In U.S. scenarios, fully remote work can reduce work-related carbon footprints by roughly 54–58%; hybrid schedules with two to four remote days per week cut footprints by about 11–29%; one remote day is only about 2%.	Climate gains exist, but they are nonlinear and depend on lifestyle and workplace configuration, not merely on using ICT.
Shi et al. [75]	In England, working from home three to five days per week yields roughly 3% less to 17% more carbon emissions than conventional work patterns depending on heated area, heating system, heating time, and indoor temperature.	Home energy can offset commuting savings; firms should avoid treating "remote" as automatically low-carbon.
Shen et al. [81]	Across 141 U.S. cities, a 1% increase in remote-work share is associated with a 1.8% reduction in daily transportation emissions per capita.	Flexible work can matter at urban scale, especially through transport rather than only individual preference.
World Resources Institute and World Business Council for Sustainable Development [82]	Scope 3 Category 7 treats employee commuting as a reportable category and allows teleworking energy to be optionally included.	Workplace design belongs inside corporate climate accounting and should be measured rather than waved at rhetorically.
International Energy Agency [76]	The IEA estimates data-centre electricity consumption at about 415 TWh in 2024 and projects around 945 TWh by 2030 in its base case.	AI-enabled flexibility has its own energy footprint; efficient models and carbon-aware deployment matter.

Appendix I. Foundation-Model Capabilities Relevant to AI-Enabled Flexibility

As mapped in Table A5, a useful distinction within the foundation-model stack is between systems that generate candidate artifacts and systems that move state across tools. The first class mainly reduces comprehension and drafting costs. The second mainly reduces coordination latency. The combination matters because many organizations are now adopting both at once.

Table A5. How foundation-model capabilities map onto the coordination frictions discussed in the main text.

Capability class	Concrete examples	Main coordination friction reduced	Main risk or limit
Multimodal capture and summarization	Meeting transcripts, action items, screenshot or slide parsing, extraction from forms and tables	Ephemeral conversation and mixed-format context become reusable artifacts	Consent, privacy, or omission of decisive nuance
Long-context reasoning	Reading long email threads, design docs, code reviews, repositories, or hours of audio or video	Less manual reconstruction of work history and rationale	Context cost, brittle synthesis, or stale information
Retrieval-augmented memory	Enterprise search over prior decisions, tickets, policies, code, and customer cases	Faster access to precedent and less duplicate work	Provenance errors, privacy leakage, or outdated documents
Copilots for writing and coding	Drafting, rewriting, translation, test generation, and code suggestions	Lower blank-page tax and faster first-pass problem solving	Overreliance, homogenization, or hidden defects
Workflow agents	Preparing agendas, updating CRM or ticketing systems, pre-filling approvals, and collecting evidence across tools	Less follow-through latency and lower process overhead	Uninspectable actions, brittle automation, or accountability gaps

Appendix J. Milestones in the Foundation-Model Stack and Their Organizational Consequences

As outlined in Table A6, the foundation-model stack matters for flexible work because each capability step removed a different coordination bottleneck. The resulting change has been cumulative: the technology first made more work *expressible*, then more work *searchable*, and now increasingly makes more work *routable*.

 doi:10.20944/preprints202604.0561.v1

Table A6. Milestones in the foundation-model stack and the specific work-design changes each one enables. Representative sources include Achiam et al. [13], Team et al. [14], Lewis et al. [15], Ning et al. [16], Yu et al. [17], Dillon et al. [37], Shao et al. [38].

Phase	Representative capability shift	What became cheaper	Resulting workstyle change
2022–2023	Conversational assistants for drafting, explanation, rewriting, and code completion	Self-service problem solving and first-pass artifact creation	Fewer routine interruptions to coworkers; more default-to-draft before asking for help
2023–2024	Workflow-embedded multimodal copilots inside email, meetings, documents, and coding tools	Turning speech, slides, screenshots, and mixed-format work into usable records	Meetings become easier to summarize and query; more work can continue asynchronously after the call
2024–2025	Long-context models plus retrieval over tickets, policies, repositories, and prior decisions	Context reconstruction and access to organizational memory	Less dependence on who happens to remember; stronger case for asynchronous handoffs across days and time zones
2025–2026	Bounded agents that gather context, prepare updates, and move state across tools with human review	Follow-through latency, routine routing, and process overhead	Fewer status pings and manual updates; more emphasis on exception handling and human approval rights
2026+	Persistent team memory, interoperable artifacts, carbon-aware model routing, and more reliable previewable multi-step workflows	Coordinating flexible work at organizational rather than individual scale	Flexible work starts to look less like an accommodation and more like an operating system, provided governance and energy use are handled well

Appendix K. Future Scenarios for Workstyle, Culture, and Work-Life Balance

As presented in Table A7, flexible work can therefore evolve into very different social equilibria even if the underlying models continue to improve.

Table A7. Three plausible medium-run equilibria for AI-enabled flexibility. The social outcome depends on mentoring design, monitoring choices, and who controls the organizational memory layer [11,57,62,63].

Scenario	Signature features	Main implication
Selective co-presence	AI is used mainly for memory, drafting, retrieval, accessibility, and bounded coordination support; on-site time is concentrated around onboarding, design sprints, conflict repair, and tightly coupled work.	This is the high-trust path. It can improve autonomy and reduce commute burdens without treating every artifact as a surveillance object.
Digital presenteeism	AI is used to accelerate response expectations, score behavior, expand documentation pressure, and preserve constant visibility even when people are off-site.	Flexibility survives in name but not in lived experience; workers gain location choice while losing boundary quality and discretion
Bifurcated flexibility	Senior or artifact-rich roles gain autonomy, while junior, operational, or place-bound roles remain highly monitored and schedule-constrained.	The main risk is not remote versus office, but a wider divide between workers with portable artifact capital and workers without it

Appendix L. Related Work and Positioning

The present paper sits at the intersection of research on AI in real work settings, foundation-model agents and their evaluation, organizational memory and knowledge visibility, and the governance of persistent digital artifacts. Its distinct claim is not simply that AI raises productivity or that agent systems are becoming more capable. Rather, it argues that the workflow-integrated foundation-model stack reprices the coordination value of distance for remote-capable knowledge work, pushing organizations toward AI-enabled flexibility with selective co-presence as the residual premium.

A first adjacent stream studies agent infrastructure rather than isolated model capability. Surveys of web agents and agent workflows emphasize tool use, orchestration, and evaluation [16,17,78]. Within this stream, He et al. [39] argues that the harness layer—tool mediation, runtime control, verification, and trace retention—is a first-class part of deployed agent capability. That framing is closely aligned with our emphasis on bounded agents and governed workflow layers, but our paper asks a different organizational question: how those layers change the default location and coordination regime of knowledge work.

A second adjacent stream focuses on measuring human–AI productivity in realistic deployment settings [36,37,40,41]. In that conversation, He et al. [69] argues that productivity claims should be reported as time-to-acceptance under explicit acceptance tests, foregrounding verification and rework rather than draft speed alone. We adopt a compatible stance, but extend it from individual-task completion to team-level handoffs, provenance, after-hours spillover, and selective co-presence.

A third adjacent stream examines governed artifacts, persistent memory, and language infrastructure. Public agent ecosystems and portfolio-style agent workflows suggest that linguistic artifacts are increasingly executable, portable, and governance-bearing rather than merely informational [83,84]. Likewise, He et al. [85] highlights a role shift from direct execution toward research direction, admissibility, and evaluation. These papers are adjacent to our notion of artifact capital, but they do not center the work-location equilibrium that is central here.

Finally, domain systems such as He et al. [86] show how provenance-first, expert-reviewed memory layers can be built in high-stakes settings. We cite this line not because the present paper is about sustainability standards, but because it offers a concrete design pattern for permissioned, inspectable organizational memory. Taken together, these adjacent papers reinforce the paper's broader

 doi:10.20944/preprints202604.0561.v1

claim that the future of flexible knowledge work depends less on generic model capability than on harnesses, governed artifacts, evaluation design, and control over the organizational memory layer.

Appendix M. Extended Literature Map

Table A8 maps the broader literature supporting the paper's core claims, grouping representative references by conceptual cluster to highlight how foundational theories connect to the paper's main arguments.

Table A8. Extended literature map for the paper's core claims.

Cluster	Representative references	Why the cluster matters for the position advanced in the main paper
Firm and coordination theory	Ronald [1], Friedrich [2], Allen [3], Garicano [5], Olson and Olson [4], Okhuysen and Bechky [48]	Establishes the baseline claim that organizations and offices historically existed partly to lower search, alignment, and monitoring costs.
Knowledge creation, transfer, and visibility	Polanyi [42], Nonaka [43], Grant [44], Argote and Ingram [45], Leonardi [60], Treem and Leonardi [61], Wang and Noe [49], Mesmer-Magnus and DeChurch [50]	Supports the central mechanism that AI changes the economics of codifying, retrieving, and reusing the traces of work without eliminating the residue of tacit knowledge.
Expertise coordination in distributed work	Faraj and Sproull [46], Lewis [47], Herbsleb and Mockus [51]	Shows that team performance depends not only on expertise itself, but on whether expertise can be located, trusted, and integrated at the right time.
Feasibility and persistence of flexible work	Dingel and Neiman [8], Mas and Pallais [21], Barrero et al. [22], Barrero et al. [23], Buckman et al. [24], Aksoy et al. [25], Aksoy et al. [26], Federal Reserve Bank of St. Louis [27], Gallup [28]	Grounds the claim that flexible work is not a temporary pandemic anomaly and that the plausible long-run equilibrium is selective flexibility rather than universal return.
Telecommuting, hybrid design, and productivity	Gajendran and Harrison [29], Allen et al. [30], Bloom et al. [31], Bloom et al. [32], Choudhury et al. [33], Angelici and Profeta [34], Gibbs et al. [35]	Supports the claim that flexible work can be productive under the right complements while preserving the important caveat that effects are heterogeneous.
Creativity, ties, mentoring, and career penalties	Yang et al. [9], Brucks and Levav [10], Emanuel et al. [11], Grund et al. [12], Leslie et al. [65], Golden and Eddleston [66], Chung and Van der Lippe [67]	Supplies the strongest reasons not to overclaim: some work genuinely benefits from co-presence, and flexibility can create visibility or promotion penalties when organizations fail to redesign evaluation.
AI in real work settings	Noy and Zhang [36], Peng et al. [87], Brynjolfsson et al. [40], Cui et al. [41], Dillon et al. [37], Dell'Acqua et al. [53], Bick et al. [18], Agrawal et al. [56], OECD/BCG/INSEAD [20], Daniotti et al. [59]	Supports the stronger claim that AI changes work patterns through drafting, search, first-pass guidance, and time reallocation, while reminding us that adoption is uneven and complementarities matter.

 doi:10.20944/preprints202604.0561.v1

Cluster	Representative references	Why the cluster matters for the position advanced in the main paper
Adoption, skill formation, and heterogeneous returns	Bick et al. [19], Cruces et al. [52], Becker et al. [88], Shen and Tamkin [68], Humlum and Vestergaard [79], Humlum and Vestergaard [89]	Clarifies why AI can simultaneously broaden participation, produce uneven returns, and leave short-run labor-market aggregates surprisingly muted.
Foundation-model stack and agents	Achiam et al. [13], Team et al. [14], Lewis et al. [15], Ning et al. [16], Yu et al. [17], Shao et al. [38], Yehudai et al. [78]	Provides the technical basis for the claim that the foundation-model stack reduces coordination frictions through multimodality, long context, retrieval, tool use, and bounded-agent workflows rather than simple text autocomplete alone.
Adjacent systems work on harnesses, governed artifacts, and evaluation	He et al. [39], He et al. [69], He et al. [83], He et al. [85], He et al. [84], He et al. [86]	Provides nearby technical and organizational framing that is complementary to, but narrower than, the present paper's focus on the work-location equilibrium of remote-capable knowledge work.
Exposure, inclusion, and unequal adoption	Eloundou et al. [90], Gmyrek et al. [91], Bloom et al. [70]	Connects flexible work to occupational exposure, disability inclusion, and the risk that AI-enabled flexibility benefits some groups much more than others.
Sustainability, climate, buildings, and ESG materiality	Marz and Şen [92], Tao et al. [80], Shi et al. [75], Shen et al. [81], World Resources Institute and World Business Council for Sustainable Development [82], Barker [71], International Energy Agency [76], Norouziasas et al. [77], Eccles et al. [72], Khan et al. [73], Friede et al. [74], He et al. [93], Zhang et al. [94], He et al. [95], He et al. [96], He et al. [97]	Shows why the environmental case is conditional and why work design now falls within sustainability accounting rather than outside it.
Governance, monitoring, and public spillovers	Kellogg et al. [62], Wood [63], Lane et al. [98], Marz and Şen [92]	Motivates the paper's governance focus: the same systems that enable portability can intensify surveillance, while broader public effects depend on institutions beyond the firm.

Appendix N. Suggested Evaluation Metrics for AI-Enabled Flexibility

The paper's normative claim is deliberately conditional, so evaluation should also be conditional. The checklist below in Table A10 is meant to make success and failure legible across productivity, sustainability, and governance dimensions rather than to optimize a single scalar metric.

 doi:10.20944/preprints202604.0561.v1

Table A10. A compact evaluation checklist for ML systems intended to support flexible work, climate claims, and governance.

Goal	Example metrics	Main failure mode to watch
Handoffs	Context-reconstruction time, unanswered clarification count, action-item carryover	Summary looks fluent but omits the decisive nuance
Institutional memory	Retrieval success on prior decisions, provenance coverage, stale-context rate	Confident retrieval of outdated or private information
Onboarding and mentoring	Ramp-up time, escalation quality, novice error rate, mentor time saved	Short-run speed gain with long-run skill loss
Worker well-being	After-hours work, task switching, quit intentions, burnout proxies	Productivity gain that simply shifts work into evenings
Carbon and resource intensity	Commute emissions, space utilization per FTE, office HVAC hours, incremental telework energy, model carbon intensity	Half-empty offices or large-model defaults make the system look flexible without actually reducing footprint
ESG and disclosure quality	Scope 3 coverage, survey response rates, assumption sensitivity, privacy complaints, appeal or override rates	Greenwashing or metric gaming that reports upside while hiding rebound effects and surveillance costs
Governance and trust	Monitoring burden, retention period, access control, appeal rate for automated judgments	Coordination tools turning into ranking or surveillance tools
Human agency in bounded-agent workflows	Fraction of automated steps previewed, override rate, rollback success, recovery time from bad actions	Convenient automation becomes unreviewable or coercive

References

1. Ronald, C. The nature of the firm. *economica* **1937**, *4*, 386–405.
2. Friedrich, H. The use of knowledge in society. *American Economic Review* **1945**, *35*, 519–530.
3. Allen, T.J. Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the R. &D. *Organization* **1977**, pp. 145–181.
4. Olson, G.M.; Olson, J.S. Distance matters. *Human–computer interaction* **2000**, *15*, 139–178.
5. Garicano, L. Hierarchies and the Organization of Knowledge in Production. *Journal of political economy* **2000**, *108*, 874–904.
6. Drucker, P. Knowledge Worker Productivity: The Biggest Challenge California Management Review **1999**.
7. Davenport, T.H. *Thinking for a living: how to get better performances and results from knowledge workers*; Harvard Business Press, 2005.
8. Dingel, J.I.; Neiman, B. How many jobs can be done at home? *Journal of public economics* **2020**, *189*, 104235.
9. Yang, L.; Holtz, D.; Jaffe, S.; Suri, S.; Sinha, S.; Weston, J.; Joyce, C.; Shah, N.; Sherman, K.; Hecht, B.; et al. The effects of remote work on collaboration among information workers. *Nature human behaviour* **2022**, *6*, 43–54.
10. Brucks, M.S.; Levav, J. Virtual communication curbs creative idea generation. *Nature* **2022**, *605*, 108–112.
11. Emanuel, N.; Harrington, E.; Pallais, A. The power of proximity to coworkers: Training for tomorrow or productivity today? Technical report, National Bureau of Economic Research, 2023.
12. Grund, C.; Harbring, C.; Klinkenberg, L. An experiment on creativity in virtual teams. *Journal of Economic Behavior & Organization* **2025**, *231*, 106926.
13. Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F.L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* **2023**.

 doi:10.20944/preprints202604.0561.v1

14. Team, G.; Georgiev, P.; Lei, V.I.; Burnell, R.; Bai, L.; Gulati, A.; Tanzer, G.; Vincent, D.; Pan, Z.; Wang, S.; et al. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530* **2024**.
15. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.t.; Rocktäschel, T.; et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems* **2020**, *33*, 9459–9474.
16. Ning, L.; Liang, Z.; Jiang, Z.; Qu, H.; Ding, Y.; Fan, W.; Wei, X.y.; Lin, S.; Liu, H.; Yu, P.S.; et al. A survey of webagents: Towards next-generation ai agents for web automation with large foundation models. In Proceedings of the Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2, 2025, pp. 6140–6150.
17. Yu, C.; Cheng, Z.; Cui, H.; Gao, Y.; Luo, Z.; Wang, Y.; Zheng, H.; Zhao, Y. A survey on agent workflow–status and future. In Proceedings of the 2025 8th International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE, 2025, pp. 770–781.
18. Bick, A.; Blandin, A.; Deming, D.J. The rapid adoption of generative AI. *Management Science* **2026**.
19. Bick, A.; Blandin, A.; Deming, D.J.; Fuchs-Schündeln, N.; Jessen, J. Mind the Gap: AI Adoption in Europe and the US. Technical report, National Bureau of Economic Research, 2026.
20. OECD/BCG/INSEAD. *The adoption of artificial intelligence in firms: New evidence for policymaking*; OECD Publishing Paris, 2025.
21. Mas, A.; Pallais, A. Valuing alternative work arrangements. *American Economic Review* **2017**, *107*, 3722–3759.
22. Barrero, J.M.; Bloom, N.; Davis, S.J. Why working from home will stick. Technical report, National Bureau of Economic Research, 2021.
23. Barrero, J.M.; Bloom, N.; Davis, S.J. The evolution of work from home. *Journal of Economic Perspectives* **2023**, *37*, 23–49.
24. Buckman, S.R.; Barrero, J.M.; Bloom, N.; Davis, S.J. Measuring work from home. Technical report, National Bureau of Economic Research, 2025.
25. Aksoy, C.G.; Barrero, J.M.; Bloom, N.; Davis, S.J.; Dolls, M.; Zarate, P. Working from home around the world. *Brookings Papers on Economic Activity* **2022**, *2022*, 281–360.
26. Aksoy, C.G.; Barrero, J.M.; Bloom, N.; Davis, S.J.; Dolls, M.; Zarate, P. The global persistence of work from home. *Proceedings of the National Academy of Sciences* **2025**, *122*, e2509892122.
27. Federal Reserve Bank of St. Louis. Work from Home Rate (WFHCOVIDMATQUESTION). FRED, Federal Reserve Bank of St. Louis, 2026. Accessed April 5, 2026.
28. Gallup. Indicators: Hybrid Work, 2025. Accessed April 5, 2026.
29. Gajendran, R.S.; Harrison, D.A. The Good, the Bad, and the Unknown About Telecommuting: Meta-Analysis of Psychological Mediators and Individual Consequences. *Journal of Applied Psychology* **2007**, *92*, 1524–1541. https://doi.org/10.1037/0021-9010.92.6.1524.
30. Allen, T.D.; Golden, T.D.; Shockley, K.M. How effective is telecommuting? Assessing the status of our scientific findings. *Psychological science in the public interest* **2015**, *16*, 40–68.
31. Bloom, N.; Liang, J.; Roberts, J.; Ying, Z.J. Does working from home work? Evidence from a Chinese experiment. *The Quarterly journal of economics* **2015**, *130*, 165–218.
32. Bloom, N.; Han, R.; Liang, J. Hybrid working from home improves retention without damaging performance. *Nature* **2024**, *630*, 920–925.
33. Choudhury, P.; Foroughi, C.; Larson, B. Work-from-anywhere: The productivity effects of geographic flexibility. *Strategic Management Journal* **2021**, *42*, 655–683.
34. Angelici, M.; Profeta, P. Smart working: work flexibility without constraints. *Management Science* **2024**, *70*, 1680–1705.
35. Gibbs, M.; Mengel, F.; Siemroth, C. Work from home and productivity: Evidence from personnel and analytics data on information technology professionals. *Journal of Political Economy Microeconomics* **2023**, *1*, 7–41.
36. Noy, S.; Zhang, W. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* **2023**, *381*, 187–192.
37. Dillon, E.W.; Jaffe, S.; Immorlica, N.; Stanton, C.T. Shifting work patterns with generative ai. Technical report, National Bureau of Economic Research, 2025.
38. Shao, Y.; Zope, H.; Jiang, Y.; Pei, J.; Nguyen, D.; Brynjolfsson, E.; Yang, D. Future of work with ai agents: Auditing automation and augmentation potential across the us workforce. *arXiv preprint arXiv:2506.06576* **2025**.

39. He, C.; Zhou, X.; Wang, D.; Xu, H.; Liu, W.; Miao, C. Harness Engineering for Language Agents: The Harness Layer as Control, Agency, and Runtime **2026**.
40. Brynjolfsson, E.; Li, D.; Raymond, L. Generative AI at work. *The Quarterly Journal of Economics* **2025**, *140*, 889–942.
41. Cui, K.Z.; Demirer, M.; Jaffe, S.; Musolff, L.; Peng, S.; Salz, T. The effects of generative AI on high-skilled work: Evidence from three field experiments with software developers. *Management Science* **2026**.
42. Polanyi, M. The tacit dimension, 1966.
43. Nonaka, I. A dynamic theory of organizational knowledge creation. *Organization science* **1994**, *5*, 14–37.
44. Grant, R.M. Toward a knowledge-based theory of the firm. *Strategic management journal* **1996**, *17*, 109–122.
45. Argote, L.; Ingram, P. Knowledge transfer: A basis for competitive advantage in firms. *Organizational behavior and human decision processes* **2000**, *82*, 150–169.
46. Faraj, S.; Sproull, L. Coordinating expertise in software development teams. *Management science* **2000**, *46*, 1554–1568.
47. Lewis, K. Measuring transactive memory systems in the field: scale development and validation. *Journal of applied psychology* **2003**, *88*, 587.
48. Okhuysen, G.A.; Bechky, B.A. 10 coordination in organizations: an integrative perspective. *Academy of Management annals* **2009**, *3*, 463–502.
49. Wang, S.; Noe, R.A. Knowledge sharing: A review and directions for future research. *Human resource management review* **2010**, *20*, 115–131.
50. Mesmer-Magnus, J.R.; DeChurch, L.A. Information sharing and team performance: a meta-analysis. *Journal of applied psychology* **2009**, *94*, 535.
51. Herbsleb, J.D.; Mockus, A. An empirical study of speed and communication in globally distributed software development. *IEEE Transactions on software engineering* **2003**, *29*, 481–494.
52. Cruces, G.; Meijide, D.F.; Galiani, S.; Gálvez, R.H.; Lombardi, M. Does Generative AI Narrow Education-Based Productivity Gaps? Evidence from a Randomized Experiment. Technical report, National Bureau of Economic Research, 2026.
53. Dell'Acqua, F.; McFowland III, E.; Mollick, E.; Lifshitz, H.; Kellogg, K.C.; Rajendran, S.; Krayer, L.; Candelon, F.; Lakhani, K.R. Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science* **2026**.
54. Bresnahan, T.F.; Brynjolfsson, E.; Hitt, L.M. Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The quarterly journal of economics* **2002**, *117*, 339–376.
55. Brynjolfsson, E.; Rock, D.; Syverson, C. The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics* **2021**, *13*, 333–372.
56. Agrawal, A.; Gans, J.S.; Goldfarb, A. Artificial intelligence adoption and system-wide change. *Journal of Economics & Management Strategy* **2024**, *33*, 327–337.
57. Gallup. The Remote Work Paradox: Higher Engagement, Lower Wellbeing, 2025. Accessed April 6, 2026.
58. Gallup. Global Data Summary — State of the Global Workplace 2025, 2025. Accessed April 6, 2026.
59. Daniotti, S.; Wachs, J.; Feng, X.; Neffke, F. Who is using AI to code? Global diffusion and impact of generative AI. *Science* **2026**, p. eadz9311.
60. Leonardi, P.M. Social media, knowledge sharing, and innovation: Toward a theory of communication visibility. *Information systems research* **2014**, *25*, 796–816.
61. Treem, J.W.; Leonardi, P.M. Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. *Annals of the International Communication Association* **2013**, *36*, 143–189.
62. Kellogg, K.C.; Valentine, M.A.; Christin, A. Algorithms at work: The new contested terrain of control. *Academy of management annals* **2020**, *14*, 366–410.
63. Wood, A.J. Algorithmic management consequences for work organisation and working conditions. Technical report, JRC working papers series on labour, education and technology, 2021.
64. Daft, R.L.; Lengel, R.H. Organizational information requirements, media richness and structural design. *Management science* **1986**, *32*, 554–571.
65. Leslie, L.M.; Manchester, C.F.; Park, T.Y.; Mehng, S.A. Flexible work practices: a source of career premiums or penalties? *Academy of Management Journal* **2012**, *55*, 1407–1428.
66. Golden, T.D.; Eddleston, K.A. Is there a price telecommuters pay? Examining the relationship between telecommuting and objective career success. *Journal of vocational behavior* **2020**, *116*, 103348.

67. Chung, H.; Van der Lippe, T. Flexible working, work–life balance, and gender equality: Introduction. *Social indicators research* **2020**, *151*, 365–381.
68. Shen, J.H.; Tamkin, A. How AI impacts skill formation. *arXiv preprint arXiv:2601.20245* **2026**.
69. He, C.; Zhou, X.; Wang, D.; Xu, H.; Liu, W.; Miao, C. Human-AI productivity claims should be reported as time-to-acceptance under explicit acceptance tests, 2026.
70. Bloom, N.; Dahl, G.B.; Rooth, D.O. Work from home and disability employment. Technical report, National Bureau of Economic Research, 2024.
71. Barker, R. Corporate sustainability reporting. *Journal of Accounting and Public Policy* **2025**, *49*, 107280.
72. Eccles, R.G.; Ioannou, I.; Serafeim, G. The impact of corporate sustainability on organizational processes and performance. *Management science* **2014**, *60*, 2835–2857.
73. Khan, M.; Serafeim, G.; Yoon, A. Corporate sustainability: First evidence on materiality. *The accounting review* **2016**, *91*, 1697–1724.
74. Friede, G.; Busch, T.; Bassen, A. ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of sustainable finance & investment* **2015**, *5*, 210–233.
75. Shi, Y.; Sorrell, S.; Foxon, T. The impact of teleworking on domestic energy use and carbon emissions: An assessment for England. *Energy and Buildings* **2023**, *287*, 112996.
76. International Energy Agency. Energy and AI. Technical report, IEA, Paris, 2025. Published 10 April 2025.
77. Norouziasas, A.; Esmaeilzadeh, A.; Yin, H.; Nord, N.; Hamdy, M. An innovative work time flexibility-based energy management system for decarbonizing office buildings. *Energy and Buildings* **2025**, *345*, 116134.
78. Yehudai, A.; Eden, L.; Li, A.; Uziel, G.; Zhao, Y.; Bar-Haim, R.; Cohan, A.; Shmueli-Scheuer, M. Survey on evaluation of llm-based agents. *arXiv preprint arXiv:2503.16416* **2025**.
79. Humlum, A.; Vestergaard, E. Large language models, small labor market effects. Technical report, National Bureau of Economic Research, 2025.
80. Tao, Y.; Yang, L.; Jaffe, S.; Amini, F.; Bergen, P.; Hecht, B.; You, F. Climate mitigation potentials of teleworking are sensitive to changes in lifestyle and workplace rather than ICT usage. *Proceedings of the National Academy of Sciences* **2023**, *120*, e2304099120.
81. Shen, S.; Wang, X.; Caros, N.; Zhao, J. The effect of remote work on urban transportation emissions: evidence from 141 cities. *Transportation Research Interdisciplinary Perspectives* **2025**, *31*, 101422.
82. World Resources Institute.; World Business Council for Sustainable Development. Scope 3 Calculation Guidance.
83. He, C.; Zhou, X.; Wang, D.; Xu, H.; Liu, W.; Miao, C. OpenClaw as Language Infrastructure: A Case-Centered Survey of a Public Agent Ecosystem in the Wild **2026**.
84. He, C.; Zhou, X.; Wang, D.; Xu, H.; Liu, W.; Miao, C. From Prompts to Portfolios: AI Agents as Agentic Multimedia Firms **2026**.
85. He, C.; Zhou, X.; Wang, D.; Xu, H.; Liu, W.; Miao, C. The AutoResearch Moment: From Experimenter to Research Director **2026**.
86. He, C.; Zhou, X.; Yu, X.; Zhang, L.; Zhang, Y.; Wu, Y.; Xiao, L.; Li, L.; Wang, D.; Xu, H.; et al. SSKG Hub: An Expert-Guided Platform for LLM-Empowered Sustainability Standards Knowledge Graphs. *arXiv preprint arXiv:2603.00669* **2026**.
87. Peng, S.; Kalliamvakou, E.; Cihon, P.; Demirer, M. The impact of ai on developer productivity: Evidence from github copilot. *arXiv preprint arXiv:2302.06590* **2023**.
88. Becker, J.; Rush, N.; Barnes, E.; Rein, D. Measuring the impact of early-2025 AI on experienced open-source developer productivity. *arXiv preprint arXiv:2507.09089* **2025**.
89. Humlum, A.; Vestergaard, E. The unequal adoption of ChatGPT exacerbates existing inequalities among workers. *Proceedings of the National Academy of Sciences* **2025**, *122*, e2414972121.
90. Eloundou, T.; Manning, S.; Mishkin, P.; Rock, D. GPTs are GPTs: Labor market impact potential of LLMs. *Science* **2024**, *384*, 1306–1308.
91. Gmyrek, P.; Berg, J.; Kamiński, K.; Konopczyński, F.; Ładna, A.; Nafradi, B.; Rosłaniec, K.; Troszyński, M. *Generative AI and jobs: A refined global index of occupational exposure*; Number 140, ILO Working Paper, 2025.
92. Marz, W.; Şen, S. Does telecommuting reduce commuting emissions? *Journal of Environmental Economics and Management* **2022**, *116*, 102746.
93. He, C.; Zhou, X.; Wu, Y.; Yu, X.; Zhang, Y.; Zhang, L.; Wang, D.; Lyu, S.; Xu, H.; Xiaoqiao, W.; et al. Esgenius: Benchmarking llms on environmental, social, and governance (esg) and sustainability knowledge. In Proceedings of the Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, 2025, pp. 14623–14664.

doi:10.20944/preprints202604.0561.v1

94. Zhang, L.; Zhou, X.; He, C.; Wang, D.; Wu, Y.; Xu, H.; Liu, W.; Miao, C. Mmesgbench: Pioneering multimodal understanding and complex reasoning benchmark for esg tasks. In Proceedings of the Proceedings of the 33rd ACM International Conference on Multimedia, 2025, pp. 12829–12836.
95. He, C.; Zhou, X.; Wang, D.; Xu, H.; Liu, W.; Miao, C. PCA-OS: A Planetary Climate Adaptation Operating System **2026**.
96. He, C.; Zhou, X.; Wang, D.; Xu, H.; Liu, W.; Miao, C. ESGlass: Glass-Box ESG and Sustainability Reports **2026**.
97. He, C.; Zhou, X.; Wang, D.; Yu, X.; Xiao, L.; Li, L.; Xu, H.; Liu, W.; Miao, C. KG4ESG: The ESG Knowledge Graph Atlas **2026**.
98. Lane, M.; Williams, M.; Broecke, S. The impact of AI on the workplace: Main findings from the OECD AI surveys of employers and workers. *OECD Social, employment and migration working papers* **2023**.