Federation Is the Way Forward for AI Agents

Herbert Woisetschläger
Technical University of Munich
Germany
h.woisetschlaeger@tum.de

Nicholas D. Lane
University of Cambridge
United Kingdom
ndl32@cam.ac.uk

Shiqiang Wang
University of Exeter
United Kingdom
s.wang9@exeter.ac.uk

Abstract

Current AI agents are largely deployed as centralized services. This architecture is prone to correlated reliability failures, unsustainable energy concentration, and structural barriers to accessing high-value private context distributed across institutions and jurisdictions. In this position paper, we argue that *federation is the way forward for AI agents*, as the most valuable context for agents is often distributed across institutions and jurisdictions. We support our position through technical, policy, and economic analysis, showing that the shift towards federated agentic architectures is both necessary and tractable. We project that federation leads to significant GDP growth potential of $783B and $470B over 10 years for the U.S. and EU, respectively. These effects are enabled by unlocking cross-boundary tasks that centralized architectures cannot reach. We also propose a deployment roadmap and research agenda that directly address the most significant federation risks.

1 Introduction

From "mainframe AI" to networked AI. In retrospect, the first era of commercial computing was defined by centralization. IBM's System/360 represented a platform breakthrough, but it was also a centralized industrial bet with a $5 billion investment (equivalent to about $53 billion today [1]), where more than 1,000 systems were ordered in its first month [32]. The next era shifted computing outward. Personal computing diffused from specialized centers to everyday endpoints. In the U.S., personal computers per 100 people rose from approximately 0.9 in 1981 to 80.8 in 2006 [59]. The core lesson is that computing capability compounds fastest when architecture evolves from centralized scarcity to distributed participation.

Today's large-model AI ecosystem resembles a new "mainframe" phase. A small number of providers host frontier models, users access those models through centralized APIs, and mission-critical workflows increasingly depend on remote inference as a single control plane. This concentration can produce short-term velocity, but it has longer term issues, including 1) correlated reliability failures when providers degrade, 2) concentration of energy and compute demand, and 3) weak access to high-value private context distributed across institutions and regions. Public incident records already reflect the first issue in practice [6], while electricity projections show the second [35].

This position paper argues that **the productivity ceiling of centralized AI agents is already visible in today's practice, and that federated agentic AI infrastructure is the way forward for a sustained and resilient productization path of AI**. The implication is not that centralized models will disappear. Instead, their architectural role should change. In a federated agentic architecture, high-capability centralized models remain important, but they are embedded in a broader network of local and domain-specific agents operating under defined trust, policy, and interoperability boundaries.

Why we focus on federation now. Federation is not a speculative concept. The initial operational proof came from federated learning, where sites train collaboratively without pooling raw data [38, 49]. Outcomes from production-like settings are now strong enough to matter for policy and systems design, including large multi-hospital deployments with measurable quality gains [16].

Preprint.

At the same time, the agentic AI paradigm has changed what must be federated. Traditional federated learning centered on sharing gradient or parameter updates. Modern agents often improve substantially through in-context adaptation, external tool use, and reasoning & action loops, even without updating model weights [11, 50, 60]. Therefore, the relevant federation target has expanded from "federated model training" to "federated agent execution," where the latter federates context access, tools, memory, policy enforcement, and orchestration across organizations. This paradigm shift comes with significant growth potential on a macro-economic level, benefiting economies with a historically high level of federation (see Figure 1, details in Section 5).

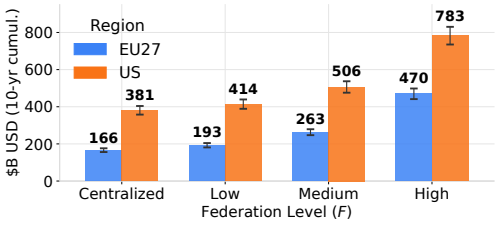

Figure 1: GDP growth potential triggered by federating agents is significant and primarily driven by task accessibility of AI agents.

2 Why Centralized Agentic Architectures Hit Limits

Centralized deployment has been a useful bootstrap strategy for the first generation of large-model products, but it does not satisfy the long-term requirements of reliability, scalability, and institutional interoperability. The limitation is architectural instead of merely operational. A single-provider control plane concentrates failure modes, power demand, and governance authority in ways that become increasingly costly as agents move from optional assistants to critical infrastructure.

Reliability becomes a correlated-risk problem. When agents depend on one remote provider for core reasoning and execution, outages are no longer isolated events. They become correlated system-wide failures. Major incidents have repeatedly surfaced in 2025 reporting, including the January 2025 ChatGPT outage [28, 58] and the June 2025 daylong disruption affecting ChatGPT, Sora, and API traffic [57, 61]. Coverage in late 2025 also reported Anthropic service disruptions affecting Claude and its console and API surfaces [15]. For low-stakes usage this is inconvenient, but for production workflows it creates a hard ceiling on trust and reliability.

Energy and compute concentration do not scale cleanly. The electricity trajectory of AI-linked digital infrastructure shows why purely centralized growth is difficult to sustain. The International Energy Agency (IEA) estimates that electricity consumption from data centers, AI, and cryptocurrency could double over four years [35]. While centralized frontier models remain essential, centralized execution cannot be the only mode.

High-value context is fragmented across trust boundaries. The most useful context for agents is private, regulated, and institution-specific, involving hospital records, enterprise logs, payment and fraud workflows, legal documents, activity traces, etc. This context is fragmented for governance reasons. The Health Insurance Portability and Accountability Act (HIPAA) imposes strict controls on protected health information [53], and General Data Protection Regulation (GDPR) sets purpose limitation and data minimization requirements with clear constraints on cross-border transfers [25]. These rules do not prohibit collaboration, but they strongly discourage raw-data centralization as the default pattern. This fragmentation reflects a broader trend towards *sovereign AI*, which is the emerging principle that institutions should control their own critical AI infrastructure and cooperate across borders without collapsing governance into a single external platform [37]. Recent European AI Factory programs and the European Health Data Space (EHDS) both reflect this direction [19, 26].

Enterprise adoption friction. In enterprise interviews and pilot discussions, several blockers appear repeatedly. First, teams are uncomfortable with full-context API exposure because it can violate internal *need-to-know* operating norms even when formal contractual protections exist. Second, buyers do not assume current large language model (LLM) SaaS pricing is the long-run steady state. Procurement and risk teams treat it as uncertain and therefore difficult to underwrite for multi-year deployment. This concern is reinforced by rapid historical price movement in frontier APIs (e.g., OpenAI's published multi-fold API price cuts and claims of order-of-magnitude declines versus earlier generations) and by external analysis arguing (in the opposite direction) that ecosystem revenue still lags infrastructure investment [13, 44, 45]. Third, vendor lock-in remains a practical governance concern. If a provider changes terms, product scope, or compliance posture, organizations can face abrupt policy misalignment after deep integration.

The above limitations of centralized architectures indicate that the next stage of agentic AI cannot be solved by larger centralized infrastructures alone. The design objective must shift from "one best model behind one service" to "many interoperable agents across many domains."

3 What "Federated" Means in the Agentic Era

The term "federated" is sometimes treated as synonymous with federated learning, but in this paper we use a broader definition that is better aligned with modern agentic systems. To avoid ambiguity, we first explain the original meaning, then define how the concept expands to agentic architectures.

3.1 Conceptualization

Federated learning, where the concept started. Federated learning is a training approach in which participating sites keep raw data local and share only selected model-side artifacts (e.g., updates or gradients) with a coordinating process. The central design idea is collaborative improvement without central raw-data pooling. Early results showed that communication-efficient optimization could reduce required communication rounds by roughly 10–100$\times$ under representative workloads [38]. Later real-world studies in healthcare further showed that such collaboration can materially improve outcomes across institutions [16]. In plain terms, being "federated" means *independent parties jointly improving intelligence while keeping sensitive data local*. That principle remains valid.

Why federated agents are broader than federated learning. In modern agentic systems, performance gains often come from mechanisms other than weight updates. LLMs now show strong few-shot and in-context adaptation behavior [11], and agentic frameworks improve capability by combining reasoning traces with external actions and tools [50, 60]. In practical deployments, this means many improvements happen through better orchestration, retrieval, memory routing, and tool invocation, even when model parameters are unchanged. As a result, "federating agents" must include more than federating model training. At minimum, it includes federating 1) context access across organizations and devices, 2) tool and API execution rights under clearly defined policy, 3) identity, trust, and permission boundaries, 4) task routing and delegation among specialized agents, 5) optional model-update exchange when it is useful and legally permissible. In this expanded view, federated learning becomes a component of a larger federated-agent stack, instead of the full definition.

Why the term "federated" should be kept. Retaining the term "federated" preserves conceptual continuity, as the field already understands federation as collaboration among autonomous participants without full central control. It is also architecturally accurate for agentic systems where multiple independent nodes (devices, enterprises, jurisdictions) coordinate under shared protocols. Moreover, it is policy-compatible language, as regulators and institutional stakeholders already reason in terms of data minimization, purpose limitation, and boundary-respecting collaboration [25, 53].

Working definition. We define a **federated agentic system** as *a network of autonomous agents that coordinate tasks, context, tools, and policies across organizational and device boundaries, with local control preserved at each node and centralized models used as optional shared services instead of single points of dependence*. This naturally includes, but is not limited to, federated learning.

3.2 Federation Across Organizations and End Users

An important design question is who participates in the federation. In practice, the answer is two-layered. Future systems need both *cross-organization federation* (institutions coordinating under policy boundaries) and *cross-end-user federation* (personal devices participating as first-class nodes).[1]

Cross-organization federation. Cross-organization federation links enterprises, public agencies, hospitals, and service providers without requiring any party to surrender full control of its data or tooling. Each organization exposes selected capabilities (e.g., retrieval endpoints, workflow tools, or verification services) under identity, permission, and audit policies. This allows agents to coordinate across domain boundaries while preserving compliance constraints. Operationally, this turns monolithic assistants into a network of specialized agents. One node may handle clinical context, another payments, another legal checks, and a frontier model may provide broad reasoning when needed. This architecture improves task quality because each domain contributes high-fidelity

[1] In federated learning literature, this is sometimes referred to as cross-silo and cross-device. We do not use those terms because organizations could also own devices and unconnected end-user devices would also be silos.

context that is difficult to replicate centrally. Figure 2 shows that mixed private and public deployment is already the dominant reality, lowering migration friction for federated execution models.

Cross-end-user federation. Cross-end-user federation extends this model to personal devices. Phones, laptops, and local assistants are not passive clients. They are decision points for identity, consent, and context release. A local agent can decide whether a request is completed on-device, delegated to an organizational agent, or escalated to a cloud model. This layer also enables global reach, as ITU reports that 78% of people aged 10+ own a mobile phone, while only 67% use the internet, and 2.6 billion people remain offline [36]. Any architecture that assumes continuous cloud access leaves out a large share of users or degrades sharply under weak connectivity.

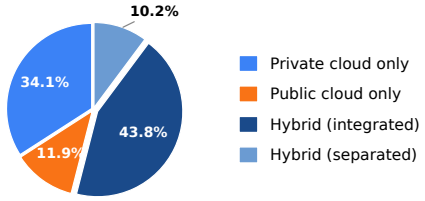

Figure 2: Hybrid topologies dominate enterprise practices, aligning with federated execution instead of single-endpoint control [42].

Agentic interfaces are shifting toward edge-first control loops. The practical interface outcome of federation is a shift from app-by-app navigation to intent-first orchestration, where traditional apps become capability providers behind an agentic user-facing front end, enabling a more conversational and less procedural interaction style while maintaining explicit permission boundaries. The most plausible near-term trajectory is a split-runtime architecture in which phones run the user-side control loop and heavier reasoning is escalated to enterprise or cloud models. On-device components handle intent capture, private context interpretation, policy checks, lightweight planning, and routing decisions, while remote tiers provide high-capability inference for difficult subproblems. Early adoption patterns support this trajectory, as OpenClaw-style assistants have shown strong user uptake through chat-first channels and lightweight onboarding paths, suggesting that users value agent access through familiar messaging surfaces before full agent-first interfaces are standardized [46].

There are trends that support this forecast. First, realistic agentic benchmarks indicate that the efficiency frontier is moving. Newer models increasingly deliver more capability per unit compute or cost, besides higher capability at larger scale. The Berkeley Function Calling Leaderboard (BFCL) evaluates tool use across non-live and live function calling, multi-turn interaction, memory, and web search [30, 47]. On this benchmark, GPT-5-mini scores 55.46% versus GPT-4.1 at 53.96%, while benchmark-estimated cost is $22.18 versus $100.75. GPT-5-nano reaches 51.45%, close to GPT-4.1, at a cost of $8.79 [30]. Figure 3 illustrates this capability-cost frontier. Cross-family results also support nontrivial efficiency gains, with Qwen3-4B at 35.68% exceeding Llama-3.3-70B at 31.90% [30]. Second, hardware capabilities for mobile inference have grown significantly. Apple introduced a Neural Engine at 38 trillion operations per second and describes this as roughly 60× faster than its first-generation baseline [9]. Apple also presents a hybrid assistant architecture in which an on-device model handles many requests and larger tasks are escalated to Private Cloud Compute [8]. While frontier models still provide a higher ceiling on difficult long-horizon tasks, the combined trend is that the continuous improvement of model efficiency and on-device inference capabilities makes phone-resident coordination increasingly viable, with enterprise or cloud tiers reserved for heavyweight reasoning.

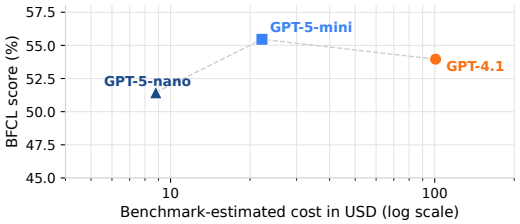

Figure 3: Capability-cost frontier on BFCL, showing newer compact variants move toward lower cost at competitive agentic performance.

Overall, organizational and end-user federation provide the coordination fabric that centralized assistants lack, with institutional specialization at the top and personal control at the edge.

4 Technical and Governance Requirements

4.1 Energy and Compute Distribution

We now discuss how compute should be distributed in face of the limits of centralized systems (Section 2). In essence, federated agentic systems should allow dynamic placement of work across device, enterprise, and cloud tiers according to energy conditions, latency targets, and policy constraints.

Table 1: Workload-placement trade-offs within a federated agentic architecture.

Dimension	In a federated system, prefer centralized execution when...	In a federated system, prefer local/distributed execution when...
Operational energy per task	workloads are homogeneous and heavily batched in efficient hyperscale facilities	workloads heterogeneous and many subtasks suitable for on-device or local execution
Network energy and traffic	task pipelines require limited back-and-forth and can be completed in a few remote calls	frequent context exchanges would otherwise create high round-trip traffic across wide-area networks
Carbon-aware placement	major data-center regions have strong low-carbon supply at execution time	tasks moved across edge/enterprise/cloud tiers to follow regional power availability in real time
Hardware lifecycle impacts	endpoint replacement cycles are stable and workload growth is mostly absorbed in shared infrastructure	local execution is paired with device longevity, repair, and reuse policies that limit e-waste growth
Resilience under outages	service continuity can tolerate provider-level concentration risk	local fallback paths are needed to avoid hard failures when cloud services degrade

Regional energy heterogeneity makes single-placement suboptimal. Electricity availability is not uniform across advanced markets. In Europe, grid coordination is explicitly cross-border. The European Union (EU) interconnection policy sets a 15% import interconnection target by 2030, and by early 2026, 16 countries were above that threshold while 9 were below the older 10% benchmark [24]. The European Network of Transmission System Operators for Electricity (ENTSO-E) governance across 40 transmission-system operators reflects this distributed operating reality [18]. In the U.S., the Lower-48 system is split into three large interconnections (Eastern, Western, ERCOT) with limited transfer capacity between them [55]. Retail demand profiles also vary sharply by state, with average annual per residential customer purchases ranging from 14,774 kWh in Louisiana to 6,178 kWh in Hawaii [56]. These differences matter for AI deployment because fixed placement choices create bottlenecks. A federated execution layer can route non-urgent or privacy-compatible workloads toward regions and tiers with better near-term power availability, while keeping latency-critical or sensitive operations local.

Federated scheduling across three compute tiers. Practically, a federated agent stack supports three complementary tiers: 1) on-device agents for intent capture, private context handling, and lightweight inference; 2) enterprise or sector agents for policy-constrained domain workflows; and 3) cloud frontier models for heavy reasoning and broad generalization. The scheduling problem is then to decide where each subtask runs based on privacy class, expected latency, model capability requirement, and current infrastructure conditions. This tiered strategy improves efficiency because not every action needs the highest-cost central model. It also improves resilience as if one tier degrades, work can often be degraded gracefully instead of failing completely.

Cost, reliability, and sustainability. When intelligence is distributed across tiers, operators gain levers that centralized stacks underuse. First, they gain *cost control* through selective escalation. Local and domain-specific agents can absorb routine tasks, reserving expensive frontier calls for high-value steps. Second, they gain *reliability control* through multi-path execution. Fallback to local or alternative agents reduces dependence on any single provider outage. Third, they gain *sustainability control* through compute placement. Flexible routing can align demand with regional energy conditions instead of concentrating all demand in a few hubs.

In essence, federation is the control model, and centralized and distributed execution are placement options selected per workload. The design target is therefore *optimal workload partitioning plus device circularity*, instead of "edge good, cloud bad." In this view, centralized compute is a critical shared resource, while federation is the control layer that decides where each step should run. Table 1 illustrates the potential control decisions *inside* a federated architecture.

4.2 Governance, Incentives, and Trust

In practice, cross-organization AI efforts fail first on governance and incentives, instead of model architecture. Federated agents therefore need institutional design that aligns 1) who controls data and tools, 2) who benefits from participation, and 3) who is accountable when coordination fails.

Governance is the bottleneck, besides model quality. In regulated domains, organizations collaborate only when control boundaries remain well-defined and auditable, which is consistent with purpose limitation in GDPR and minimum-necessary handling in HIPAA [25, 53]. This implies a governance baseline that each delegation decision must be attributable to a principal, each cross-boundary action must be policy-checked, and each output must carry provenance metadata sufficient

for audit and dispute resolution. Being *auditable* requires a tamper-evident execution trail identifying which agent acted, which policy checks were applied, what data was accessed, and what outputs were produced. Therefore, policy cannot be a thin wrapper around model calls. It must be part of runtime execution with identity-scoped tool access, consent and revocation paths, and enforceable separation between local memory and shared outputs. Without these controls, "federation" degrades into ad hoc API chaining with unclear responsibility.

Incentive alignment determines participation. Participation drivers differ by sector. In healthcare, collaboration is motivated by public value and scientific reputation, especially when rare cohorts are too sparse at any single institution [16]. In finance, data is a competitive asset, so participation requires clear return of investment (ROI) such as reducing cross-institution fraud losses. Federation strategy should therefore be sector-specific, with incentive design as a first-class systems requirement.

Trust must be operationalized across boundaries. At minimum, participating organizations need 1) verifiable identity for each agent node, 2) least-privilege execution for tools and connectors, 3) tamper-evident logs for delegated actions, and 4) fallback behavior when upstream agents are unavailable. High-capability centralized models remain valuable, but should operate as optional reasoning services within a governed federated system and not as the sole trust anchor.

Access control and safety protocols are a prerequisite for federation. Federation multiplies the number of trust boundaries that agents must cross, making access control and inter-agent safety protocols a first-class design requirement instead of an afterthought. At minimum, a federated deployment requires policy-enforced capability scoping, verifiable agent identity, tamper-evident audit trails, and composable safety constraints that remain valid across multi-hop delegation chains. Governments begin to recognize this stark need and have released significant funding programs to enable research on scalable and trustworthy agent coordination [5, 17, 22, 41]. This emerging cross-Atlantic policy consensus signals that the security and access-control layer of federated agent systems requires dedicated public investment before ecosystem-scale deployment becomes viable.

First movers and coalition effects. Federation requires first movers willing to absorb integration risk before network effects emerge. Medical research is a natural candidate, supported by mandates such as the National Institutes of Health (NIH) Data Management and Sharing Policy [40], the International Committee of Medical Journal Editors (ICMJE) data-sharing requirements [34], and the EHDS Regulation [26]. Early technical coalitions already exist, including multi-hospital federated deployments [16] and the MELLODDY consortium across ten pharmaceutical companies [23, 43]. Adjacent sectors can reuse these templates with different incentive contracts and risk controls. For regions with decentralized data landscapes, competitive advantage can come from better federation protocols and governance models instead of larger centralized training runs, making governance design a core source of system performance, legitimacy, and long-run scalability.

5 Economic Rationale

To quantify the stakes of the architectural choice, we translate the effects of federation into projected GDP impact. In short, we can show using well-accepted economic models that moving from fully centralized AI to a high federation level projects up to $783B and $470B in additional 10-year cumulative GDP growth for the U.S. and EU27, respectively (roughly doubling the centralized baseline in both regions, see Figure 1).

We derive these numbers from the total factor productivity (TFP) growth, which captures economy-wide efficiency gains more precisely than GDP alone, since GDP conflates productivity with factor accumulation. The task-based framework from economics research [2, 3] delivers TFP estimates directly from observable task-level cost savings, which we then translate to GDP via $\Delta\text{GDP} = \text{GDP} \times \text{TFP}_{10\text{yr}} \times 1.85$. We estimate the economy-wide benefit is roughly $1.85\times$ the direct benefit, meaning their ripple effects through investment and growth is about $1.85\times$ of the initial impact. Such an estimation is a common practice in economics research [4]. The formalism and a more extensive discussion on the economic model we use is provided in Appendix A. The core TFP result is that a high federation level increases cumulative 10-year TFP growth by 0.75 percentage points for the U.S. and by 0.85 percentage points for the EU27.

The EU27 gains relatively more because its regulatory landscape blocks more tasks from centralized access in the first place (a $2.8\times$ amplification versus $2.1\times$ for the U.S.). In the following, we explain how each federation effect drives these numbers.

Mapping federation effects to economic output. We use a task-based model of the economy composed of a large number of tasks where where each one is routed either to automated capital (e.g., AI) or human labor [2, 4]. AI improves aggregate output through four channels, and increasing federation amplifies each channel to a different degree. Each task in the pipeline produces output differently depending on who or what handles it. For automated tasks, the output depends on capital availability, how well AI fits the task (a task-capital complementarity factor), and resource utilization costs. For human-performed tasks, output depends on labor input and how well human skills match the task. Federation shifts both the complementarity factors and the utilization costs across the board.

Hulten's theorem [31] acts as an aggregation layer. It lets us sum up the per-task efficiency gains into a single economy-wide productivity change measured by TFP. Each task's contribution is weighted by its share of total economic output, multiplied by the unit cost savings AI enables for that task. The key architectural assumption is that centralized and federated AI do not just perform the same tasks more or less efficiently. Instead, they operate on fundamentally different task sets. Federation unlocks a strictly larger set of automatable tasks. This means that federation does not just tune the parameters of the production function. It expands the function's domain.

Channel 1 – New Tasks: the biggest leverage. These are tasks that require cross-boundary data access and have no centralized substitute (e.g., cross-border clinical reasoning, fraud detection across banks, or supply-chain optimization across factories). No single provider can perform a cross-border diagnostic reasoning task if the input data cannot legally leave its jurisdiction. Federation addresses this by keeping data in place and coordinating only policy-filtered outputs.

This channel is the dominant driver of the EU27 federation premium. GDPR, HIPAA, and frameworks such as the EHDS block a large share of high-value cross-boundary tasks from any centralized provider, giving the EU27 a much larger pool of unlockable tasks than the U.S. To give a sense of scale, the EHDS alone projects approximately €11B (approx. $12B at 2024 exchange rates[2]) in savings from cross-border health-data reuse [20]. That is a single regulated domain. As an estimate, if even 3–5% of currently inaccessible cross-boundary tasks became feasible through federation, the implied TFP contribution could outgrow the centralized baseline in data-fragmented regions [2].

Channel 2 – Task Enrichment: richer context improves existing tasks. Even for tasks that centralized systems can already perform, federation improves output quality by giving agents access to richer cross-boundary context. For example, a triage agent that can query anonymized summary statistics from multiple partner hospitals produces better recommendations than one limited to its own institution's data. Existing centralized AI deployments provide a lower bound on the productivity gains achievable within a single organization. That is, a 14% average productivity increase with generative AI (34% for novice workers) [12] and up to 55.8% faster task completion with AI pair programming [48] are possible. These studies measure only within-organization single-provider deployments. The additional productivity increment from cross-boundary federated context has no direct empirical estimate yet, which is currently a limitation. Federation extends this channel into domains where centralized architectures face institutional boundaries that limit data access.

Channel 3 – Deepening of Tasks: multi-provider routing reduces lock-in and capital cost. Federation also lowers the cost of automated tasks by enabling multi-provider routing. Instead of being committed to one vendor's inference stack, an orchestrating agent can select the cheapest provider at runtime. Ofcom survey data show that 43.2% of cloud users cite time and cost of change as a switching barrier [42], and 35.3% considered switching but did not, citing portability difficulties (31.5%) and staff retraining costs (32.8%). Federation treats provider selection as a runtime scheduling decision instead of a long-term architectural commitment, reducing friction for already-automated tasks.

Channel 4 – Automation: marginal benefit at trust boundaries. Automation (shifting tasks from labor to capital) can be achieved by both centralized and federated AI agents. The federation-specific gain is that tasks requiring compliance verification across organizational boundaries can shift from human to automated execution under a federated delegation model (e.g., a factory agent supervised by a headquarters compliance agent). This is the weakest channel, because the incremental TFP contribution is modest relative to Channels 1–3, which is reflected in the small gap between centralized and low-federation scenarios described in Appendix A.

Putting the numbers together. For the U.S., the centralized baseline of 0.71% cumulative 10-year TFP growth (derived from [2]) grows to 0.77% at low federation, 0.94% at medium, and 1.46% at

[2]We use the 2024 GDP base in our calculation, thus using the 2024 exchange rate here.

high. In the EU, the baseline of 0.46% grows to 0.54%, 0.73%, and 1.31%, respectively, reflecting the IMF estimate that EU regulation puts over 30% of AI-related gains at risk [39]. The gap between the U.S. and EU27 federation multipliers reflects the general principle that *federation matters most where centralization is structurally impossible, instead of merely inconvenient.* To formalize this, we consider the federation design objective shared across both regions, specifically minimizing the *risk-adjusted* cost of success, defined as $C_{\text{success}} = C_{\text{exec}} + C_{\text{coord}} + p_{\text{outage}} \times L_{\text{outage}} + p_{\text{compliance}} \times L_{\text{compliance}}$, where C, p, and L stand for cost, failure probability, and loss, respectively, with the subscripts denoting their specific subject matters. All these parameters are in principle observable from deployment data. For example, provider outage histories and GDPR fine databases [14, 33] provide a useful overview of outage or compliance risks, respectively. In practice, however, the dominant source of uncertainty is the coordination cost C_{coord}, for which no empirical estimates exist yet.

Federation should therefore be evaluated as a staged infrastructure investment. Upfront costs concentrate in integration and control-plane capabilities, while returns accrue across several measurable dimensions. These returns can be grounded empirically in the number of tasks unlocked through collaboration (e.g., via EHDS option value), labor productivity gains [12, 48], avoided outage losses $p_{\text{outage}} \times L_{\text{outage}}$ from production incident data [29], and user adoption trends from surveys such as Ofcom data in Channel 3.

Result. Taking the 2024 GDP base of the EU ($19.4T) and the U.S. ($29T), we can then translate the potential TFP gains discussed along the four channels directly into GDP growth projections (Figure 1).[3] A more detailed description on the calculation is in Appendix B.

6 Roadmap for Federated Agents

To realize the full growth potential, we propose a deployment path that executes in four phases: 1) *intra-organization hybrid*, which establishes local fallback behavior and auditable tool permissions under a single governance boundary; 2) *bilateral federation*, which validates policy-compliant cross-boundary collaboration without raw-data pooling; 3) *public-user onboarding*, which introduces opt-in consumer users once on-device inference, latency, and safety thresholds are met; and 4) *multi-party federation*, which targets ecosystem-level interoperability with standardized delegation semantics and graceful degradation. Each phase builds a specific capability and generates observable evidence before committing to the next. A phase should proceed only when leading indicators clear predefined thresholds, including indicators such as payback period per phase, change in cost per completed workflow, outage-adjusted productivity, escalation-rate reduction to premium models, and compliance-incident cost avoided. We provide more details on the four phases in Appendix C.

The deployment path requires a parallel research agenda. Deployment without research risks lock-in to fragile interfaces and research without deployment risks benchmarks that do not reflect real governance, outage, and incentive constraints. The unifying problem is preserving user and institutional control while maintaining reliable capability across heterogeneous runtimes, policies, and providers. The main research themes, challenges, and key research questions are detailed in Table 2.

Table 2: Research agenda for federated agents

Theme	**Challenge**	**Key research questions**
Delegation contracts and typed capabilities	Delegation crosses organizations with different trust and execution assumptions.	What minimal type system can encode actions, data classes, and side effects? How capabilities should narrow/expire during long plans? How delegated execution can be proven contract-compliant?
Cross-boundary policy composition	Locally valid decisions can become globally invalid after multi-hop delegation.	How to compose conflicting policies, compute the strictest valid plan automatically, and preserve auditable policy provenance across hops?
Context orchestration and residency	Task context must be assembled across device, enterprise, and cloud without violating residency constraints.	Which context elements must remain local versus delegable? How can cross-tier retrieval and transformation avoid raw sensitive-data exposure? How does memory partitioning support reproducible context reconstruction?

[3] We note that our calculations represent gross productivity gains and do not net out upfront federation infrastructure investment. Thus, the gains shown in Figure 1 should be treated as an upper bound. We leave the analysis of transition costs for future empirical work.

Outage-aware multi-provider planning	Reliability depends on graceful degradation under provider failure.	How to pre-compile fallback plans, reroute under policy constraints, and measure reliability gains versus coordination overhead in multi-provider execution?
Attribution, settlement, and liability	Multi-party execution chains blur value creation and responsibility.	How to attribute marginal value per delegation hop, settle compute and data costs across asymmetric participants, and assign liability when harm emerges from a chain instead of one endpoint?
Cross-tier configuration portability and user retention	Users face hidden behavior drift when skills, memory, and policies move across runtimes, and may disengage if complexity reduces efficiency.	How to make skills, memory schemas, and safety preferences portable across device, organization, and cloud tiers? Which invariants must hold across model and provider changes? Which complexity thresholds predict abandonment by non-expert users?

7 Alternative Views

Federation is not the only viable architectural direction, especially since extensive regulation and the need for significant investments into building the basis of federated agent architectures can stifle the growth potential. Beyond the key points in this section, we provide more details in Appendix D.

Complexity. The strongest objection is that a centralized assistant is easier to implement, benchmark, and operate. This is valid in the short term. The practical response is sequencing and scope control. Federated systems should begin with narrow delegation patterns and fallback rules, then widen only when reliability and governance evidence is strong.

Quality variance. In a federated workflow, one weak node can reduce end-to-end performance even when other nodes are strong. This creates a risk of inconsistent user experience. Mitigation requires capability-aware routing and verifiable delegation contracts so tasks are assigned by demonstrated competence instead of static trust assumptions.

Model behavior variation. Even with stable prompts, tools, and infrastructure, provider-side model updates or runtime changes can alter outputs and action choices in ways that break downstream assumptions. This motivates operational controls beyond prompt engineering, including version pinning windows, behavioral regression suites, and conformance checks for high-risk workflows.

Policy drift. Even when each institution is internally compliant, the composed workflow may violate a cross-boundary constraint through unintended delegation paths. This is why policy composition and auditability are important design requirements. Federation fails if policy is evaluated locally but never validated globally.

Security surface. More nodes, tools, and connectors create more opportunities for compromise. Thus, the architecture must default to least privilege, scoped credentials, and tamper-evident execution trails, with isolation between personal context, institutional context, and shared task artifacts. Federation is safer than centralization only when trust boundaries are engineered as first-class runtime objects.

8 Conclusion

Federation is the way forward for AI agents. Centralized frontier models remain essential, but reliable and governable agentic systems require a federated control plane coordinating across devices, organizations, and shared model infrastructure. The projected GDP gains are driven primarily by cross-boundary tasks that centralized architectures cannot legally or technically reach, concentrated in regulated sectors such as healthcare, finance, legal services, and manufacturing where data-localization rules have historically been an obstacle towards AI adoption. Significant public investment in secure and trustworthy multi-agent coordination, including from ARIA, EU Horizon Europe, NSF, and DARPA, signals that security, governance, and multi-agent coordination is increasingly recognized as a foundational research priority in the context of AI agents. Realizing these gains requires solving governance before scaling architecture. Delegation must be attributable, with policy-checked cross-boundary actions and traceable outputs. Coordination costs, quality variance under policy filtering, and the absence of empirical estimates for the new-tasks channel remain the principal open risks. Researchers and builders should prioritize delegation protocols, cross-boundary policy composition, and accountable multi-party execution now, before the ecosystem defaults to proprietary lock-in. The window to establish open interoperable federation standards is narrow, as architectural defaults set today will determine which organizations and which regions can participate in the next generation of AI-driven economic activity.

References

[1] Value of 1964 US dollars today. Inflation Tool. URL https://www.inflationtool.com/us-dollar/1964-to-present-value.

[2] Daron Acemoglu. The simple macroeconomics of AI. *Economic Policy*, 40(121):13–58, 2025.

[3] Daron Acemoglu and Pascual Restrepo. The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6): 1488–1542, 2018.

[4] Daron Acemoglu and Pascual Restrepo. Tasks, automation, and the rise in US wage inequality. *Econometrica*, 90(5):1973–2016, 2022.

[5] Advanced Research and Invention Agency. Scaling trust programme thesis, 2026. URL https://aria.org.uk/media/dkhlumky/scaling-trust-programme-thesis.pdf.

[6] Anthropic. Claude status history. https://status.claude.com/history.

[7] Tom M. Apostol. *Calculus, Volume 2: Multi-Variable Calculus and Linear Algebra with Applications to Differential Equations and Probability*. John Wiley & Sons, New York, 2nd edition, 1969.

[8] Apple. Introducing apple intelligence, the personal intelligence system that puts powerful generative models at the core of iPhone, iPad, and Mac. Apple Newsroom, June 2024. URL https://www.apple.com/newsroom/2024/06/introducing-apple-intelligence-for-iphone-ipad-and-mac/.

[9] Apple. Apple introduces M4 chip. Apple Newsroom, May 2024. URL https://www.apple.com/newsroom/2024/05/apple-introduces-m4-chip/.

[10] David Rezza Baqaee and Emmanuel Farhi. The macroeconomic impact of microeconomic shocks: Beyond Hulten's theorem. *Econometrica*, 87(4):1155–1203, 2019.

[11] Tom Brown, Benjamin Mann, Nick Ryder, et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901, 2020.

[12] Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. Generative AI at work. *The Quarterly Journal of Economics*, 140(2):889–942, 05 2025.

[13] David Cahn. AI's $600b question. Sequoia Capital perspective, June 2024. URL https://www.sequoiacap.com/article/ais-600b-question/.

[14] CMS Law. GDPR enforcement tracker. URL https://www.enforcementtracker.com/.

[15] Dominic-Madori Davis. Anthropic reports outages, Claude and Console impacted. TechCrunch, September 2025. URL https://techcrunch.com/2025/09/10/anthropic-reports-outages-claude-and-console-impacted/.

[16] Ittai Dayan, Holger R. Roth, Aoxiao Zhong, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*, 27:1735–1743, 2021.

[17] Defense Advanced Research Projects Agency. Information innovation office (I2O) office-wide broad agency announcement, 2025. URL https://sam.gov/opp/091b4d199d7241dbbb04b8d36eb88a16/view.

[18] ENTSO-E. About entso-e. ENTSO-E website. URL https://www.entsoe.eu/about/.

[19] EuroHPC Joint Undertaking. AI factories. Official program page. URL https://www.eurohpc-ju.europa.eu/ai-factories_en.

[20] European Commission. European health data space regulation (EHDS) – public health portal. DG SANTE webpage, . URL https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en.

[21] European Commission. ESCO: European skills, competences, qualifications and occupations. https://esco.ec.europa.eu, . Accessed 2025.

[22] European Commission. EU invests over €307 million into artificial intelligence and related technologies, 2026. URL https://digital-strategy.ec.europa.eu/en/news/eu-invests-over-eu307-million-artificial-intelligence-and-related-technologies.

[23] European Commission CORDIS. MELLODDY: Machine learning ledger orchestration for drug discovery. Horizon 2020 project fact sheet, 2024. URL https://cordis.europa.eu/project/id/831472.

[24] European Commission, Directorate-General for Energy. Electricity interconnection targets. European Commission website. URL https://energy.ec.europa.eu/topics/infrastructure/electricity-interconnection-targets_en.

[25] European Union. Regulation (EU) 2016/679 (GDPR). Official Journal of the European Union, L119, 2016. URL https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[26] European Union. Regulation (EU) 2025/327 on the European health data space. Official Journal of the European Union, 2025. URL http://data.europa.eu/eli/reg/2025/327/oj.

[27] Eurostat. Structure of earnings survey, 2022. URL https://ec.europa.eu/eurostat/web/microdata/structure-of-earnings-survey.

[28] Tom Gerken. ChatGPT back online after outage which hit thousands worldwide. BBC News, January 2025. URL https://www.bbc.com/news/articles/c30d80lg579o.

[29] GitHub. GitHub status incidents history. URL https://www.githubstatus.com/history.

[30] Gorilla LLM Team, UC Berkeley. Berkeley function calling leaderboard (BFCL) v4 – overall results. URL https://gorilla.cs.berkeley.edu/leaderboard.html.

[31] Charles R. Hulten. Growth accounting with intermediate inputs. *Review of Economic Studies*, 45(3):511–518, 1978.

[32] IBM. The IBM system/360. IBM Heritage. URL https://www.ibm.com/history/system-360.

[33] INPLP. GDPR fines database. International Network of Privacy Law Professionals. URL https://gdpr-fines.inplp.com/.

[34] International Committee of Medical Journal Editors. Recommendations: Clinical trials (registration and data sharing). ICMJE Recommendations. URL https://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html.

[35] International Energy Agency. Electricity 2024: Analysis and forecast to 2026. Technical report, IEA, 2024. URL https://www.iea.org/reports/electricity-2024.

[36] International Telecommunication Union. New global connectivity data shows growth, but divides persist, 2023. URL https://www.itu.int/en/mediacentre/Pages/PR-2023-11-27-facts-and-figures-measuring-digital-development.aspx.

[37] McKinsey & Co. What is sovereign AI?, 2026. URL https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-sovereign-ai.

[38] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1273–1282, 2017.

[39] Florian Misch, Ben Park, Carlo Pizzinelli, and Galen Sher. Ai and Productivity in Europe. *IMF Working Papers*, 2025(067):1, 4 2025. ISSN 1018-5941. doi: 10.5089/9798229006057.001. URL http://dx.doi.org/10.5089/9798229006057.001.

[40] National Institutes of Health. NOT-OD-21-013: Final NIH policy for data management and sharing. NIH Notice, October 2020. URL https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html.

[41] National Science Foundation. Advancing artificial intelligence (AI) agent ecosystems through the NSF pathways to enable secure open-source ecosystems (PESOSE) program, 2026. URL https://www.nsf.gov/funding/opportunities/dcl-advancing-artificial-intelligence-ai-agent-ecosystems.

[42] Ofcom and Context Consulting. Cloud services market research data tables. Ofcom cloud services market study supporting dataset, 2023. URL available from main page at https://www.ofcom.org.uk/internet-based-services/cloud-services/cloud-services-market-study.

[43] Martijn Oldenhof, Gergely Ács, Balázs Pejó, et al. Industry-scale orchestrated federated learning for drug discovery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[44] OpenAI. New models and developer products announced at DevDay. OpenAI product announcement, November 2023. URL https://openai.com/index/new-models-and-developer-products-announced-at-devday/.

[45] OpenAI. GPT-4o mini: Advancing cost-efficient intelligence. OpenAI release note, July 2024. URL https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/.

[46] openclaw. OpenClaw repository: Personal AI assistant gateway. URL https://github.com/openclaw/openclaw.

[47] Shishir G Patil, Huanzhi Mao, Fanjia Yan, Charlie Cheng-Jie Ji, Vishnu Suresh, Ion Stoica, and Joseph E. Gonzalez. The berkeley function calling leaderboard (BFCL): From tool use to agentic evaluation of large language models. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pages 48371–48392. PMLR, 2025.

[48] Sida Peng, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. The impact of AI on developer productivity: Evidence from GitHub Copilot. *arXiv preprint arXiv:2302.06590*, 2023. URL https://arxiv.org/abs/2302.06590.

[49] Nicola Rieke, Jonny Hancox, Wenqi Li, et al. The future of digital health with federated learning. *npj Digital Medicine*, 3:119, 2020.

[50] Noah Shinn, Federico Cassano, Ashwin Gopinath, et al. Reflexion: language agents with verbal reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 36, pages 8634–8652, 2023.

[51] Maja S. Svanberg, Wensu Li, Martin Fleming, Brian C. Goehring, and Neil C. Thompson. Beyond AI exposure: Which tasks are cost-effective to automate with computer vision? 2024.

[52] U.S. Bureau of Labor Statistics. Occupational employment and wage statistics (OEWS), 2024. URL https://www.bls.gov/oes/.

[53] U.S. Department of Health and Human Services. Summary of the HIPAA privacy rule. HHS Health Information Privacy. URL https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html.

[54] U.S. Department of Labor. O*NET OnLine: Occupational information network. https://www.onetonline.org. Accessed 2025.

[55] U.S. Energy Information Administration. U.S. electric system is made up of interconnections and balancing authorities. Today in Energy, 2016. URL https://www.eia.gov/todayinenergy/detail.php?id=27152.

[56] U.S. Energy Information Administration. How much electricity does an american home use?, 2024. URL https://www.eia.gov/tools/faqs/faq.php?id=97&t=3.

[57] Jess Weatherbed. ChatGPT's daylong outage is nearly fixed. The Verge, June 2025. URL https://www.theverge.com/news/684141/openai-chatgpt-sora-outage-issues-june-2025.

[58] Kyle Wiggers. ChatGPT suffered a major outage this morning, but OpenAI says it is back up. TechCrunch, January 2025. URL https://techcrunch.com/2025/01/23/chatgpt-suffered-a-major-outage-this-morning-but-openai-says-its-back-up/.

[59] World Bank. Personal computers (per 100 people), indicator IT.CMP.PCMP.P2. World Development Indicators API. URL https://api.worldbank.org/v2/country/USA/indicator/IT.CMP.PCMP.P2?format=json&per_page=20000.

[60] Shunyu Yao, Jeffrey Zhao, Dian Yu, et al. React: Synergizing reasoning and acting in language models. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=WE_vluYUL-X.

[61] Maxwell Zeff. ChatGPT is having a partial outage. TechCrunch, June 2025. URL https://techcrunch.com/2025/06/10/chatgpt-is-having-a-partial-outage/.

Appendix

A Production-Function Model Details

This appendix provides the formal details underlying the economic model in Section 5. We follow the task-based framework [2, 3] and extend it with a federation index $F \in [0, 1]$ that captures how many cross-boundary tasks have been unlocked.

A.1 Baseline Framework

The economy produces a single final good Y from a set of tasks $\mathcal{T}_F \subseteq \mathbb{R}_+$, where $\mathcal{T}_F$ denotes the feasible task set at federation level F and tasks are indexed by $z \in \mathcal{T}_F$. Tasks are *gross complements* with substitution elasticity $\sigma < 1$, i.e., no task can be freely replaced by expanding another. This complementarity is the key structural feature, i.e., a bottleneck in any single task drags down the total output, so expanding the task set (increasing the cardinality $|\mathcal{T}_F|$ of the set $\mathcal{T}_F$) can have disproportionately large macroeconomic effects even when the new tasks start with a small GDP share.

The aggregate output is

$$Y(F) = \left(\int_{\mathcal{T}_F} y(z, F)^{\frac{\sigma-1}{\sigma}} \, dz \right)^{\frac{\sigma}{\sigma-1}}, \quad \sigma < 1, \tag{1}$$

and each task is assigned either to automated capital or human labor:

$$y(z, F) = \begin{cases} A_K \, \gamma_K(z, F) \, k(z) \,/\, \rho(z, F) & \text{if } z \le I(F) \ \text{ (automated)}, \\ A_L \, \gamma_L(z, F) \, l(z) & \text{if } z > I(F) \ \text{ (labor-performed)}. \end{cases} \tag{2}$$

Variables and parameters. The mathematical notations in the above equations are explained as follows. $I(F)$ is the automation threshold; tasks at or below it are automated, and we require $I(F) \le |\mathcal{T}_F|$ so that the threshold lies within the feasible task set. The task-level capital and labor productivity are $\gamma_K(z, F)$ and $\gamma_L(z, F)$, respectively. The per-unit capital cost for task z is $\rho(z, F)$. A_K and A_L are economy-wide productivity shifters. The total factor endowments $\bar{L}$ and $\bar{K}$ are fixed. Competitive markets allocate $l(z)$ and $k(z)$ across tasks so that $\int_{\{z \in \mathcal{T}_F \,:\, z > I(F)\}} l(z) \, dz = \bar{L}$ and $\int_{\{z \in \mathcal{T}_F \,:\, z \le I(F)\}} k(z) \, dz = \bar{K}$.

TFP and Hulten's theorem. Under competitive markets and constant returns to scale, the first-order effect of any parameter change on aggregate TFP can be expressed as a cost-share-weighted integral of task-level cost savings [31]:

$$\frac{\Delta \text{TFP}}{\text{TFP}} = \int_{\mathcal{T}_F} w_z(F) \cdot s_z(F) \, dz, \tag{3}$$

where $w_z(F)$ is task z's share of GDP and $s_z(F) = (v_z(F) - c_z(F)) \,/\, c_z(F)$ is the unit cost saving (v_z is the unit value of the task's output, c_z is its unit cost). A federated task contributes positively ($s_z > 0$) if and only if federation lowers its effective production cost.

A.2 How the Federation Index F Enters Each Parameter

The key structural assumption of our extension is that *centralized and federated AI operate on different task sets*, with $|\mathcal{T}_0| \ll |\mathcal{T}_1|$. We parameterize each production function input by F as follows.

Task set, $\mathcal{T}_F$. At $F = 0$, only tasks whose required data resides within one organizational boundary are feasible. As F increases, tasks requiring cross-boundary data access become executable. Formally,

$$\mathcal{T}_F = \mathcal{T}_{\text{cent}} \cup \Delta\mathcal{T}_F, \qquad \mathcal{T}_{\text{cent}} \cap \Delta\mathcal{T}_F = \emptyset,$$

where $\mathcal{T}_{\text{cent}}$ is the fixed set of tasks on the centralized frontier and $\Delta\mathcal{T}_F$ is the set of tasks that federation additionally unlocks. In jurisdictions with strong data-localization rules (e.g., GDPR, HIPAA, or EHDS), $|\Delta\mathcal{T}_F|$ is large. In markets with fewer restrictions, it is small. This asymmetry is the primary reason the EU27 relatively benefits more from federation than the U.S. in our calculation.

Automation threshold, $I(F)$. At $F = 0$, only tasks whose compliance can be verified within one organization can be automated. Federation enables supervisory delegation across trust boundaries, shifting additional tasks from labor to capital (e.g., a factory agent supervised by a compliance agent at headquarters). $I(F)$ increases with F but is bounded above by $|\mathcal{T}_F|$, the largest index in the current feasible task set.

Labor productivity, $\gamma_L(z, F)$ and A_L. At $F = 0$, a labor-performed task draws only on locally available context. As F increases, agents can access richer cross-boundary context through policy-filtered summaries, federated queries, and delegated verification. This raises $\gamma_L(z, F)$ for tasks that benefit from distributed information (e.g., diagnostic reasoning, fraud triage, regulatory compliance). The empirical calibration point for this channel comes from centralized AI productivity studies, where Brynjolfsson et al. [12] found a 14% average productivity increase in customer service deployments (34% for novice workers), and Peng et al. [48] found up to 55.8% faster task completion with AI pair programming. A limitation though is that these studies measure only within-organization single-provider deployments. The additional increment $\gamma_L(z, F) - \gamma_L(z, 0)$ for tasks that genuinely require cross-boundary data has no direct empirical estimate yet. We treat the centralized numbers (Table 3) as lower bounds and flag the cross-boundary increment as the main open calibration gap.

Capital productivity and cost, $\gamma_K(z, F)$ and $\rho(z, F)$. At $F = 0$, automated tasks rely on a single provider's context and compute. Federation allows automated tasks to query verification or data-summary agents at partner organizations, improving output quality (γ_K increases). It also enables multi-provider routing at runtime, which reduces per-unit capital cost (ρ decreases) by dissolving vendor lock-in. Ofcom survey data show that 43.2% of cloud users cite switching cost as a barrier [42], confirming the empirical relevance of this channel.

Labor and capital allocation, $l(z)$ and $k(z)$. Total endowments $\bar{L}$ and $\bar{K}$ are fixed, but *competitive market clearing* reallocates them as F changes. When $\mathcal{T}_F$ expands, both labor and capital spread over a larger task set. When I shifts upward, workers move from newly automated tasks to remaining labor-performed tasks, and capital expands into the newly automated zone. The full general-equilibrium reallocation is captured through the spending weights $w_z(F)$ and cost savings $s_z(F)$ in (3).

A.3 Equilibrium and Comparative Statics

The model admits a competitive equilibrium in which wages clear the labor market and rental rates clear the capital market. The key comparative static is $\partial Y/\partial F$. Because Y depends on F both through the integrand and through the upper limit of integration, we parameterize the task set as the interval $\mathcal{T}_F = [0, b(F)]$ with $b(F)$ being a differentiable scalar function of F, so that Leibniz's rule applies directly [7]. Differentiating (1) with respect to F then gives:

$$\begin{aligned}\frac{\partial Y}{\partial F} = \underbrace{\Phi(F) \cdot y(b(F), F)^{\frac{\sigma-1}{\sigma}} \cdot b'(F)}_{\text{new tasks (Leibniz boundary term)}} + \underbrace{\frac{\partial Y}{\partial I} \cdot \frac{\partial I}{\partial F}}_{\text{automation}} + \underbrace{\int_{\mathcal{T}_F} \frac{\partial Y}{\partial \gamma_L(z)} \cdot \frac{\partial \gamma_L(z)}{\partial F}\, dz}_{\text{labor complementarity}} \\ + \underbrace{\int_{\mathcal{T}_F} \frac{\partial Y}{\partial \gamma_K(z)} \cdot \frac{\partial \gamma_K(z)}{\partial F}\, dz}_{\text{capital quality}} + \underbrace{\int_{\mathcal{T}_F} \frac{\partial Y}{\partial \rho(z)} \cdot \frac{\partial \rho(z)}{\partial F}\, dz}_{\text{capital cost}},\end{aligned} \tag{4}$$

where $\Phi(F) = \frac{\sigma}{1-\sigma}\left(\int_{\mathcal{T}_F} y(z, F)^{\frac{\sigma-1}{\sigma}} dz\right)^{-\frac{1}{1-\sigma}} > 0$ is the outer derivative of the Constant Elasticity of Substitution (CES) aggregator evaluated at the current task set, and $b'(F) = \partial b/\partial F \geq 0$ is the rate at which the task frontier expands with federation. The first term is the Leibniz boundary contribution, which means that as F increases and new tasks enter at the frontier $b(F)$, the aggregate output rises by the marginal CES value of the task at the boundary, scaled by the rate of task-frontier expansion $b'(F)$. The remaining four terms capture within-task parameter changes integrated over the current feasible set, following the comparative statics structure of Acemoglu and Restrepo [3, 4]. Each term is non-negative under Assumption 3 in Appendix A.7, so $\partial Y/\partial F \geq 0$ as long as federation does not actively degrade any parameter. The relative magnitude of each term differs by region. In data-fragmented jurisdictions like the EU27, the Leibniz boundary term dominates because $b'(F)$ is large. In the U.S., $\partial\rho/\partial F$ is relatively more important because multi-provider routing reduces lock-in even where centralization is viable.

A.4 Net Benefit and the Viability Threshold

In Section 5, we introduce a risk-adjusted cost of success:

$$C_{\text{success}}(F) = C_{\text{exec}} + C_{\text{coord}}(F) + p_{\text{outage}} \times L_{\text{outage}} + p_{\text{compliance}} \times L_{\text{compliance}}. \tag{5}$$

Here, C_{exec} is the fixed per-workflow execution cost (compute, inference, and operational overhead) that is independent of federation level. $C_{\text{coord}}(F)$ is the coordination overhead that grows with F

(integration, policy negotiation, and identity management). The probability $p_{\text{outage}} \in [0,1]$ is the per-period probability of a provider outage and $L_{\text{outage}} \geq 0$ is the expected output loss conditional on such an outage. The probability $p_{\text{compliance}} \in [0,1]$ is the per-period probability of a compliance incident and $L_{\text{compliance}} \geq 0$ is the associated expected penalty or remediation cost. The net benefit ratio is $\mathcal{B}(F) = Y(F)/C_{\text{success}}(F)$. Federation is economically justified when $\mathcal{B}(F)$ is increasing, i.e., $\partial\mathcal{B}/\partial F > 0$. Applying the quotient rule and noting that $C_{\text{success}}(F)^2 > 0$, this reduces to

$$\frac{\partial Y}{\partial F} \cdot C_{\text{success}}(F) \;-\; Y(F) \cdot \frac{\partial C_{\text{success}}}{\partial F} > 0.$$

Three structural observations support $\mathcal{B}(F)$ being increasing above a minimum viability threshold. First, $\partial Y/\partial F$ operates through $\mathcal{T}_F$, which creates net-new productive activities instead of redistributing gains from existing tasks. Second, $C_{\text{coord}}(F)$ (i.e., the dominant term in $\partial C_{\text{success}}/\partial F$) likely grows sub-linearly, since protocol standards (delegation schemes, policy languages, identity frameworks) carry high fixed costs but near-zero marginal costs once adopted, analogous to the standardization dynamics of internet protocols and API ecosystems. Third, the outage and compliance cost components are both decreasing in F (since multi-path execution reduces p_{outage} and better policy enforcement reduces $p_{\text{compliance}}$), while C_{exec} is flat by assumption. Together these imply $\partial C_{\text{success}}/\partial F$ may be negative once coordination overhead has been amortized.

Note, the viability threshold is an open empirical question. The minimum F at which $\mathcal{B}(F)$ is increasing, i.e., where output gains begin to outpace the marginal cost of coordination, has yet to be studied. It depends on 1) the regulatory environment, which determines how many tasks are blocked at $F = 0$; 2) the share of tasks that benefit from cross-boundary context, which determines $|\Delta\mathcal{T}_F|$; and 3) the fixed cost of federation infrastructure, which determines how quickly C_{coord} amortizes.

A.5 Why the Centralized Baseline Is Not the Ceiling

A natural objection is that centralized AI could eventually reach the same outputs as federation, given enough time and capability. The model shows why this is not the case for a large class of tasks. Tasks in $\Delta\mathcal{T}_F$ are not slower or less capable versions of centralized tasks. They are tasks whose *data inputs* cannot legally or contractually be pooled into a single location (due to GDPR, HIPAA, trade-secret protections, or competitive data moats). No amount of centralized model capability can perform a cross-border diagnostic reasoning task if the patient records cannot legally leave their origin jurisdiction. Federation makes these tasks feasible by coordinating only policy-filtered outputs and never raw data. The production function is therefore structurally different at $F = 0$ and $F > 0$: the feasible task set itself changes ($|\mathcal{T}_0| \ll |\mathcal{T}_1|$).

A.6 Identification and Calibration

Several model parameters can be grounded in existing data.

Task sets ($\mathcal{T}_F$, $\mathcal{T}_{\text{cent}}$). These can be proxied using industry occupation-task mappings (e.g., ESCO or O*NET) [3, 4, 21, 54], then estimating what fraction of tasks requires cross-boundary data. The IMF estimates that over 30% of potential AI productivity gains in Europe are at risk from data-localization regulation [39], providing an upper bound on $|\Delta\mathcal{T}_F|$ for the EU27.

Spending weights ($w_z(F)$). These map to observable GDP shares by occupation. Acemoglu [2] calibrates U.S. weights from U.S. Bureau of Labor Statistics employment and wage data [52]. Analogous constructions are feasible for the EU27 using Eurostat structure-of-earnings surveys [27].

Productivity parameters ($\gamma_L(z,0)$, $\gamma_K(z,0)$). Lower bounds come from existing centralized AI deployment studies [12, 48]. The federated increment $\gamma_L(z,F) - \gamma_L(z,0)$ for cross-boundary tasks has no direct estimate yet and is the main open empirical gap in the model.

Cost parameters (C_{success}). Outage probabilities p_{outage} can be estimated from provider status histories. Compliance costs can be estimated from GDPR fine databases [14, 33]. Coordination costs $C_{\text{coord}}(F)$ are the hardest to calibrate ex ante but can be tracked as federated deployments mature.

A.7 Assumptions and Limitations

1. **Gross complementarity ($\sigma < 1$).** Tasks cannot be freely substituted. This is supported by the empirical observation that removing any single occupation from an economy causes disproportionate output loss [3, 4]. If $\sigma \geq 1$ (gross substitutes), the new-tasks channel

($\Delta\mathcal{T}_F$) would have negligible macroeconomic impact because existing tasks would simply expand to compensate.

2. **Competitive markets and constant returns to scale.** The Hulten decomposition in (3) is exact under these conditions. Under imperfect competition, it remains a first-order approximation that becomes exact only under constant markups [10, 31]. The direction of the bias depends on how markups are distributed across federated versus centralized tasks.
3. **Federation does not reduce task quality.** We assume $\gamma_L(z, F) \geq \gamma_L(z, 0)$ and $\gamma_K(z, F) \geq \gamma_K(z, 0)$ for all $z \in \mathcal{T}_0$ and $F \geq 0$. This could fail if federation introduces communication latency, model heterogeneity, or policy-induced information loss. $C_{\text{coord}}(F)$ partially accounts for this, but a fully general model would allow γ to decrease for some tasks. Appendix D discusses when and why this assumption may not hold.
4. **Coordination cost is monotone and concave in F.** We assume $C_{\text{coord}}(F)$ grows with F but at a decreasing rate as protocols standardize. If coordination costs grow super-linearly instead (e.g., due to combinatorial policy negotiation), $\mathcal{B}(F)$ could turn negative at moderate F before standardization reduces marginal cost.
5. **Centralized baseline.** The TFP calibrations from Acemoglu [2] and Misch et al. [39] represent $F \approx 0$. Any federation-driven gain is additive on top of these baselines, not a replacement for them.
6. **Static comparative statics.** The model compares equilibria at different F levels without modeling transition dynamics. Moving from $F = 0$ to $F > 0$ requires protocol development, identity frameworks, and institutional trust-building. The cost side is partially captured by $C_{\text{coord}}(F)$, but the time path and path-dependence of adoption are not modeled.
7. **Exclusion of capital expenditure for federation**. The model compares equilibria at different federation levels F without accounting for the upfront capital investment required to reach a given F. Building delegation infrastructure, deploying identity frameworks, and achieving protocol standardization all constitute stock investments that precede the productivity gains captured in ΔGDP. While $C_{coord}(F)$ partially proxies for ongoing coordination overhead as a flow cost, it does not represent the full capital expenditure profile of federation adoption. As a result, the net benefit ratio $\mathcal{B}(F) = Y(F)/C_{success}(F)$ and the derived GDP figures should be interpreted as upper bounds on the true net gain. The gap between gross and net gains depends on the size of federation infrastructure investment and the discount rate applied to the transition period, neither of which is empirically estimated. Quantifying this gap is an important direction for future work, and is partially tractable once federated deployments as described in Appendix C begin generating observable data.
8. **No distributional effects.** The model does not capture labor displacement or inequality. Federation increases $I(F)$ (automating some tasks) while expanding $\mathcal{T}_F$ (creating others), but the displaced workers and the beneficiaries of new tasks may not be the same people. These distributional consequences are outside the scope of the aggregate TFP analysis.

B GDP Calculation from TFP

We now discuss how the 10-year cumulative GDP gains in the paper are computed from the TFP estimates.

GDP baseline. We use U.S. GDP of \$29,000B and EU27 GDP of €17,900B converted at the 2024 average rate of €1 = \$1.082, giving a EU27 GDP of \$19,368B. These are the nominal baselines against which TFP growth rates are applied.

From TFP to GDP: Hulten's theorem and the GDP multiplier. The task-based framework delivers a GDP impact equal to the task-level cost savings weighted by each task's GDP share [2]. Aggregating across all tasks gives a TFP change, and by Hulten's theorem this is the first-order approximation to the GDP change holding factor prices fixed. A second effect arises from the capital-deepening response. When TFP rises, investment increases, and total GDP growth exceeds TFP growth by a factor of approximately $1.85\times$, derived from the full general-equilibrium framework of Acemoglu and Restrepo [4] with a U.S. capital share of approximately 27%,

$$\Delta\text{GDP}_{\text{10yr}} = \text{GDP}_{\text{baseline}} \times \text{TFP}_{\text{10yr}} \times 1.85. \tag{6}$$

We apply the formula separately for the U.S. and EU27 using their respective baselines (Table 3). The multiplier of $1.85\times$ reflects the midpoint of Acemoglu's two scenarios where 1) capital rises proportionally with TFP, giving a multiplier of approximately $1.4\times$, and 2) full investment response from Acemoglu and Restrepo [4] yielding approximately 2.1–$2.4\times$ [2]; specifically, $1.85\times$ is the midpoint of $1.4\times$ and $2.3\times$, the latter being the midpoint of the 2.1–$2.4\times$ range.

We further use a Monte Carlo simulation (n = 2,000 draws, uniform parameter sampling over the ranges in Table 4) to propagate uncertainty in the key calibration parameters cross-boundary productivity increment λ_{xb} (i.e., by how much productivity grows if agents can collaborate across institutional borders) and the unit cost savings from automation s_{auto} to the TFP and GDP estimates. The centralized baseline has no uncertainty (it is the Acemoglu/IMF calibration).

Parameter ranges. Table 4 uses the following uniform parameter ranges, sourced from the cited literature and the Monte Carlo sweep. We provide an overview how the quantified parameters map to practical considerations in Table 5.

Role of λ_{xb} and s_{auto}. The two stochastic parameters in the Monte Carlo sweep enter the TFP calculation as follows. The cross-boundary productivity increment $\lambda_{\text{xb}} \in [0, 1]$ scales the labor productivity gain on tasks that require data from multiple institutions. Concretely, for any task $z \in \Delta\mathcal{T}_F$, the federated labor productivity is:

$$\gamma_L(z, F) = \gamma_L(z, 0) \cdot (1 + \lambda_{\text{xb}} \cdot F), \tag{7}$$

so that $\lambda_{\text{xb}} = 0$ recovers the centralized baseline and $\lambda_{\text{xb}} = 1$ at $F = 1$ doubles task-level labor productivity at full federation. The automation share $s_{\text{auto}} \in [0, 1]$ is the fraction of currently labor-performed tasks that become automatable under a federated delegation model, i.e., the share of tasks that shift from $z > I(F)$ to $z \leq I(F)$ as F increases. It enters the automation threshold as:

$$I(F) = I(0) + s_{\text{auto}} \cdot F \cdot \big(|\mathcal{T}_F| - I(0)\big), \tag{8}$$

so that at $F = 1$ a share s_{auto} of the remaining labor-performed tasks has been automated. The U.S. and EU27 values of s_{auto} are calibrated from Acemoglu [2] and Misch et al. [39] respectively, while λ_{xb} has no direct empirical estimate and is treated as the primary uncertainty parameter in our study in the Monte Carlo sweep, consistent with the open calibration gap noted in Appendix A.7.

- **Centralized** ($F = 0.0$): U.S. TFP = 0.710% = Acemoglu [2] central estimate; EU27 TFP = 0.464% = IMF central estimate adjusted downward to reflect the IMF's finding that EU regulation puts over 30% of potential AI productivity gains at risk [39].
- **Low** ($F = 0.25$): U.S. $\lambda_{\text{xb}} \in [0.10, 0.25]$, EU27 $\lambda_{\text{xb}} \in [0.25, 0.40]$, U.S. $s_{\text{auto}} = 0.23$, EU27 $s_{\text{auto}} = 0.16$.
- **Medium** ($F = 0.50$, 10th-90th percentile sweep): U.S. $\lambda_{\text{xb}} \in [0.10, 0.25]$, EU27 $\lambda_{\text{xb}} \in [0.25, 0.55]$, U.S. $s_{\text{auto}} = 0.23$, EU27 $s_{\text{auto}} = 0.16$.
- **High** ($F = 1.00$): U.S. $\lambda_{\text{xb}} \in [0.10, 0.25]$, EU27 $\lambda_{\text{xb}} \in [0.25, 0.55]$, U.S. $s_{\text{auto}} = 0.23$, EU27 $s_{\text{auto}} = 0.16$.

The ranges for λ_{xb} are constructed from indirect evidence rather than direct calibration, as no empirical estimate for cross-boundary agentic productivity increments currently exists (Appendix A.7, Assumption 7).

The lower bound of 0.10 reflects the conservative position that cross-boundary context access yields at least a modest increment above the within-organization productivity gains documented in centralized deployments [12, 48]. Setting $\lambda_{\text{xb}} = 0$ would imply that federation adds nothing beyond single-institution deployment, which is inconsistent with evidence from multi-institutional federated learning studies. In the most directly comparable empirical setting, Dayan et al. [16] find that a federated model trained across 20 institutions achieves a 16% improvement in predictive performance and a 38% gain in generalizability relative to single-institution models, suggesting that cross-boundary collaboration materially improves output quality in regulated domains.

The upper bounds diverge by region for structural reasons. For the U.S., the upper bound of 0.25 reflects the more limited fragmentation of the domestic task landscape, where many high-value tasks are already accessible to centralized providers. For the EU27, the upper bound of 0.55 reflects the larger pool of tasks blocked from centralized access by data-localization regulation. Misch et al. [39] estimate that over 30% of potential AI productivity gains in Europe are at risk from such regulation, implying that cross-boundary access would unlock a correspondingly larger productivity increment when those barriers are lifted through federation. Both upper bounds remain below the 38% generalizability gain reported by Dayan et al. [16], providing a conservative ceiling consistent with the available evidence. We treat model-quality improvements as a proxy for task-level output quality gains, consistent with the interpretation of $\gamma_L(z, F)$ as a productivity shifter rather than a pure accuracy measure. This is the best available estimate, as no empirical studies on cross-boundary agentic productivity increments have yet been conducted.

The U.S. value $s_{\text{auto}} = 0.23$ is taken directly from Acemoglu [2], who calibrates it from the share of AI-exposed tasks that Svanberg et al. [51] find profitable to automate within a 10-year horizon. The EU27 value $s_{\text{auto}} = 0.16$ is derived from Misch et al. [39], who apply the same framework to European labor markets and adjust downward to reflect regulatory constraints on AI adoption across tasks and sectors.

Table 3: Cumulative 10-year TFP and GDP estimates by federation level and region. GDP computed with U.S. GDP = $29,000B and EU27 GDP = $19,368B.

Federation Level	U.S. TFP (%)	EU27 TFP (%)	U.S. ΔGDP ($B)	EU27 ΔGDP ($B)
Centralized	0.710	0.464	381	166
Low	0.772	0.538	414	193
Medium	0.944	0.734	506	263
High	1.459	1.311	783	470

Table 4: TFP 90% confidence intervals and derived GDP ranges for stochastic federation levels. Centralized baseline has no CI (fixed Acemoglu/IMF calibration).

Federation Level	U.S. TFP 90% CI	EU27 TFP 90% CI	U.S. ΔGDP 90% CI ($B)	EU27 ΔGDP 90% CI ($B)
Centralized	—	—	—	—
Low	[0.686, 0.879]	[0.490, 0.600]	[368, 472]	[176, 215]
Medium	[0.847, 1.137]	[0.640, 0.860]	[454, 610]	[229, 308]
High	[1.287, 1.737]	[1.020, 1.580]	[690, 932]	[365, 566]

Table 5: Effects of federated AI agents on the main channels of macro-economic production.

Channel	**Low** ($F \approx 0.3$)	**Medium** ($F \approx 0.6$)	**High** ($F \approx 1$)
New tasks (N)	Intra-org cross-department tasks: hospital chain routes patient summaries across units under one DPA. Modest N increase constrained by org boundary.	Bilateral cross-org tasks: two hospitals in different member states coordinate under EHDS exemptions. Meaningful N increase into regulated domains.	Multi-party ecosystem tasks: pan-EU pharmacovigilance, cross-border fraud detection across dozens of agents. Large N expansion into domains centralization cannot reach.
Complementarities (γ_L, A_L)	Multi-provider fallback: agents switch models on outage. Gains limited to reliability.	Cross-org context routing: triage agent draws on partner-institution summaries with consent. γ_L improves because existing tasks produce better outputs.	Full cross-boundary context: agents access policy-filtered context from dozens of orgs routinely. Maximum γ_L gains into regulated domains.
Deepening (γ_K, ρ)	Vendor diversification: tasks route to cheapest capable provider, reducing ρ. γ_K gains minimal.	Scoped external context: automated subtasks query verification agents at partner orgs, improving γ_K. Delegation protocols reduce ρ further.	Rich multi-source context: every automated task draws on distributed policy-filtered data. Maximum γ_K, minimum ρ via commodity routing.
Automation (I)	Limited new automation: single-provider tasks already automated. Federation adds redundancy only.	Supervised boundary automation: compliance-supervised agents automate tasks at trust boundaries (e.g., factory OT supervised by HQ). I increases modestly.	Full boundary automation: any task whose compliance can be verified through federated delegation can be automated. Maximum I expansion.

C Deployment Path

Federation is best approached as staged infrastructure investment instead of a single architectural commitment. The phases below are ordered by governance complexity and federation scope. Each phase assumes the controls established in the previous one are stable and auditable.

Deployment path. The deployment path is structured along four phases:

Phase 1 (intra-organization hybrid). Start with a single organization operating a three-tier stack (device, enterprise, cloud). The objective is controlled reliability and policy enforcement under one governance boundary including local fallback behavior, escalation rules, and auditable tool permissions.

Phase 2 (bilateral federation). Add one external organizational partner and formalize cross-boundary contracts for identity, delegation scope, logging, and revocation. The objective is to validate that policy-compliant collaboration improves task quality without requiring raw-data pooling.

Phase 3 (public-user onboarding). Introduce opt-in consumer users after minimum readiness thresholds are met for on-device inference capability, latency, battery impact, and safety controls. Public-facing deployment should begin with personal and local workflows (for example scheduling, personal knowledge, and app/tool orchestration), while sensitive enterprise actions remain gated behind authorization and policy checks.

Phase 4 (multi-party federation). Expand to a sector-level network with heterogeneous participants and model providers, including public-user-facing services at scale. The objective is interoperability at ecosystem scale supported by portable trust policies, standardized delegation semantics, and graceful degradation under partial outages.

D Alternative Views: Detailed Analysis

This appendix expands the gaps and risks identified in Section 7.

The new-tasks channel is empirically unobserved. The model's strongest claim is $\partial|\mathcal{T}_F|/\partial F > 0$, meaning federation unlocks productive tasks that centralization cannot reach. This is structurally plausible when data cannot legally leave its jurisdiction. No federated-agent deployment at macroeconomic scale has been observed yet. The EHDS €11B (~$12B at 2024 exchange rates) projection is a policy estimate, not an ex-post measurement. The productivity studies cited in Section 5 (14% average gains from centralized AI [12], up to 55.8% faster task completion [48]) measure only within-organization, centralized access. The federated increment on cross-boundary tasks ($\gamma_L(z, F) - \gamma_L(z, 0)$ for tasks requiring data from multiple institutions) has no direct estimate. The entire economic case therefore rests on a channel whose magnitude is theoretically motivated but not yet empirically grounded.

Gross complementarity may not hold at task level. The macroeconomic significance of new tasks depends on $\sigma < 1$ (gross complementarity). If tasks are gross substitutes ($\sigma \geq 1$), expanding $\mathcal{T}_F$ by admitting tasks in $\Delta\mathcal{T}_F$ has negligible TFP impact because existing tasks simply expand to compensate. The task-based literature assumes $\sigma < 1$ based on occupational evidence, but this is an aggregate property. At the level of individual federated-agent tasks, substitution elasticities may be higher. A cross-border fraud detection task that federation enables may substitute for an existing centralized fraud screening task instead of complementing it. If many new federated tasks are substitutes for existing centralized ones, the TFP contribution of tasks in $\Delta\mathcal{T}_F$ is smaller than the aggregate model suggests.

Coordination costs may grow superlinearly. The model assumes $C_{\text{coord}}(F)$ increases with F but at a decreasing rate, reflecting protocol standardization and fixed-cost amortization. This may not hold. Cross-boundary policy negotiation involves combinatorial complexity. Each new institutional participant adds not one policy edge but a set of bilateral policy pairs that must be composed. For delegation chains of length ℓ across k institutions each with m policy constraints, the number of compositions that must be checked grows as $O(k^\ell m^\ell)$, and no single party may have designed or be able to control the full composed policy. Our cost function captures this through $C_{\text{coord}}(F)$, but if coordination costs grow super-linearly, the net benefit $\mathcal{B}(F)$ could turn negative at moderate F levels before standardization reduces marginal cost. The viability threshold is then harder to reach than the concave cost model implies.

Federation may degrade task quality for some tasks. The model assumes $\gamma_L(z, F) \geq \gamma_L(z, 0)$ and $\gamma_K(z, F) \geq \gamma_K(z, 0)$ for all $z \in \mathcal{T}_0$. This could fail in at least three ways. Communication latency across trust boundaries can reduce response speed below interactive-use thresholds. Policy filtering strips information from context, and the stripped context may be insufficient for the task. Heterogeneous model providers produce inconsistent outputs that break downstream assumptions. Our cost function captures some of this through $C_{\text{coord}}(F)$, but an honest accounting should acknowledge that for some tasks, federation may actively reduce quality relative to a single-provider baseline. The net TFP effect then depends on whether the quality gains from newly feasible tasks ($\Delta\mathcal{T}_F$) outweigh quality losses on existing tasks.

The marginal-gains objection holds where centralization is viable. Where centralized data access is legally and technically tractable, federation may yield only incremental improvements. The U.S. domestic market, with a single legal framework, large cloud incumbents, and limited data-localization requirements, is exactly this case. Coordination costs and transition frictions may eat most of the productivity benefit, leaving federation only slightly better than centralization for many tasks. The economic case for federation is strongest where centralization is structurally impossible, not where it is merely inconvenient. This means the model predicts heterogeneous adoption. The EU27, with its data-fragmentation barriers, should adopt federation earlier and more extensively than the U.S., not because federation is universally superior, but because the centralized alternative cannot operate across the regulatory boundaries that define most high-value tasks.

The model ignores transition dynamics. Moving from $F = 0$ to $F > 0$ requires investment in protocols, identity frameworks, policy languages, and institutional trust. The production function compares equilibria at different F levels but does not model the transition path. If the transition takes a decade, the net present value of federation gains depends heavily on discount rates. If early federated deployments produce negative experiences (quality variance, security incidents, policy

conflicts), adoption could stall before the threshold where $\mathcal{B}(F)$ turns positive. The static model also does not account for path dependence. Early architectural choices, such as which delegation protocol or policy language becomes dominant, may lock in coordination costs or prevent later improvements.

Labor displacement and distributional effects are unmodeled. Federation shifts the automation threshold $I(F)$ upward, automating tasks that were previously labor-performed. It also creates new tasks in $\Delta\mathcal{T}_F$, but the workers displaced from automated tasks may not be the same workers who benefit from new ones. The resulting distributional effects are not captured by the aggregate TFP analysis. If the gains from federation accrue primarily to capital owners and the costs fall on displaced workers, the political economy of adoption becomes significantly harder, even if the aggregate $\mathcal{B}(F)$ is positive.

Security and policy composition remain unsolved. Federation increases the number of nodes, tools, connectors, and delegation hops. This expands the attack surface and creates new failure modes. Each cross-boundary delegation introduces a potential confidentiality or integrity violation. Even when each institution is internally compliant, composed workflows may violate cross-boundary constraints through unintended delegation paths. Policy composition (composing GDPR purpose limitation with institution-specific retention rules, for example) is not a solved problem. Federation is safer than centralization only when trust boundaries are engineered as first-class runtime objects, and this engineering work is substantial.