Chapter 11 / Capítulo 11

SAP Superintelligence Series
ISSN: 3144-5624
ISBN (complete work): 978-9915-9851-0-7

Superintelligence Series Volume 4: One Person Company: Artificial Intelligence, Autonomous Entrepreneurship and the Future of Business
ISBN (this volume): 978-9915-9966-3-9
DOI: 10.62486/978-9915-9966-3-9.Ch11
Pages: 118-128

Check for updates

© 2026 The authors. This is an open access book chapter distributed under the terms of the Creative Commons Attribution (CC BY) 4.0 License.

Multi-agent artificial intelligence systems as the workforce of the one-person company

Sistemas multiagente de inteligencia artificial como fuerza de trabajo de la empresa unipersonal

Alfredo Javier Pérez Gamboa[1]

1 Centro de Investigación en Educación, Naturaleza, Cultura e Innovación para la Amazonia – CIENCIA. Colombia.

ABSTRACT

The one-person company, a firm in which a single founder is supported by artificial intelligence rather than by employees, depends on a specific technical claim, that a collection of autonomous agents can perform the work of a workforce. This narrative review synthesizes the architectures, the business functions and the failure modes of multi-agent artificial intelligence systems to assess that claim. The synthesis finds that multi-agent systems have moved from single large language models to orchestrated collections of specialized agents, that these collections now perform bounded business functions such as process automation, procurement, marketing and workflow orchestration, and that their failure modes, error propagation, collusion, security gaps and a missing oversight regime, bound the reliance that a founder can place on them. The review concludes that the multi-agent workforce is a real and rapidly maturing technical substrate, that its business functions are demonstrated for bounded workflows rather than for open-ended judgment, and that the solopreneur who adopts it inherits a set of oversight obligations that the technology does not discharge.

Keywords: multi-agent systems; agentic artificial intelligence; large language model agents; one-person company; autonomous business operations; orchestration

RESUMEN

La empresa unipersonal, una firma en la que un único fundador se apoya en la inteligencia artificial en lugar de en empleados, depende de una afirmación técnica concreta, que un conjunto de agentes autónomos puede realizar el trabajo de una plantilla. Esta revisión narrativa sintetiza las arquitecturas, las funciones de negocio y los modos de fallo de los sistemas multiagente de inteligencia artificial para evaluar esa afirmación. La síntesis encuentra que los sistemas multiagente han pasado de modelos de lenguaje únicos a colecciones orquestadas de agentes especializados, que esas colecciones realizan ya funciones de negocio acotadas como la automatización de procesos, las compras, el marketing y la orquestación de flujos de trabajo, y que sus modos de fallo, la propagación de errores, la colusión, las brechas de seguridad y un régimen de supervisión ausente, limitan la confianza que un fundador puede depositar en ellos. La revisión concluye que la plantilla multiagente es un sustrato técnico real y que madura con rapidez, que sus funciones de negocio están demostradas para flujos acotados y no para el juicio abierto, y que el solopreneur que la adopta hereda un conjunto de obligaciones de supervisión que la tecnología no descarga.

Palabras clave: sistemas multiagente; inteligencia artificial agéntica; agentes de modelos de lenguaje; empresa unipersonal; operaciones autónomas de negocio; orquestación

INTRODUCTION

The one-person company, a firm in which a single founder is supported by artificial intelligence rather than by a payroll, rests on a technical claim that has only recently become testable, that a collection of autonomous software agents can perform the work that employees, managers and contractors once performed. The claim has moved from the margins to the mainstream of research in a short period, and the growth is measurable, with the survey and framework literature on large language model agents expanding from a handful of definitions in 2023 to a large body of surveys, benchmarks and deployment studies by 2026.[1,2] The term used throughout this chapter is the multi-agent system, defined as a set of large language model agents that decompose a goal into roles, messages and actions and coordinate among themselves to complete it, and the alternative names of the one-person company, solopreneur and super-individual are retained as keywords. The scale of the shift is reflected in the taxonomy, which moved from a single autonomous agent to a spectrum of single-agent, multi-agent and orchestrated systems in the space of three years, and the taxonomies that order the field are themselves a measure of how fast it grew.[1,3] The workforce claim is therefore no longer futuristic but present, and the question it raises is no longer whether the agents exist but whether they can be trusted, which is the question the failure-mode literature addresses and the one-person company cannot avoid. The chapter therefore treats the multi-agent system as a hypothesis about a workforce, and it tests the hypothesis against the three bodies of evidence rather than against the claim that the practitioner commentary makes.

The state of the art is no longer a single model answering a prompt but a division of labour, in which specialized agents assume the roles of a software engineer, a marketer, a procurement officer or a controller and exchange messages through an orchestration layer.[4,5] The architectures that support this division have been classified and compared, and the comparison shows a common core of role assignment, shared memory and message passing beneath a diversity of coordination schemes, from debate to role-playing to centralized orchestration.[1,6] The business applications have followed the architectures, and the evidence now reports agentic systems performing process automation, marketing, procurement and workflow orchestration in small and medium enterprises, with the bibliometric evidence mapping a pattern of rapid growth concentrated in that segment.[7,8] The safety and oversight questions have followed the applications, with a parallel literature documenting error propagation, collusion, security gaps and the absence of a settled evaluation regime, and the critical assessments conclude that the capability is real and the oversight is lagging.[9,10] The transition from the single model to the multi-agent system was reported as the decisive step, because the single model answers and the multi-agent system acts, and the difference between answering and acting is the difference between a tool and a workforce.(1,4)

The gap is a synthesis that treats the multi-agent system specifically as the workforce of the one-person company, and that joins the architecture, the business function and the failure mode into a single assessment of the solopreneur's claim. The existing surveys cover the architectures, the enterprise studies cover the business functions, and the safety studies cover the failure modes, but the three literatures have not been read together against the question that the one-person company poses, which is whether a founder can rely on agents as a workforce rather than as a tool. The gap is consequential because the one-person company is presented in practitioner commentary as a near-term outcome, and a founder or a funder who accepts that presentation without the failure-mode evidence accepts a claim that the literature does not yet support. The three literatures each answer part of the question, and none answers the whole of it, because the architect does not measure the business function, the business study does not measure the failure, and the safety study does not measure the firm.

The objective of this review is to synthesize the architectures, the business functions and the failure modes of multi-agent artificial intelligence systems as the workforce of the one-person company, and to delimit the conditions under which that workforce is a reliable substitute for a payroll. The chapter asks a single question, whether the multi-agent system is a workforce or a set of tools wearing the name of a workforce, and the answer is assembled from the three literatures rather than asserted from one. The answer is expected to be conditional rather than categorical, because the evidence reviewed below shows a substrate that is real at one margin and absent at another.

https://doi.org/10.62486/978-9915-9966-3-9.Ch11

DEVELOPMENT

MULTI-AGENT ARCHITECTURES AND THE DIVISION OF LABOUR

The first result concerns the architecture of the multi-agent system, and the main datum is that the field has standardized on a division of labour in which specialized agents, not a single model, carry the work. It was observed across the survey literature that a multi-agent system decomposes a goal into roles, assigns each role an agent with its own memory and instruction, and coordinates the agents through message passing, and this decomposition was reported as the defining feature that distinguishes a multi-agent system from a single large language model.[1,4] The number of coordination schemes was found to exceed the number of shared primitives, with debate, role-playing, centralized orchestration and consensus all resting on the same three primitives of roles, memory and messages.[2,6] The architecture surveys reported that the field had reached a degree of standardization sufficient to support comparison, with the deployment of agents classified by their autonomy, their action space and their memory, and the lifecycle survey added a graph augmentation that places the workflow, not the model, at the centre of the system.[5,11] The role of the persona was recorded as a further architectural variable, with cognitive personality priors shaping how an agent reasons about a task, and the recursive self-improvement literature reported that agents can be composed into pipelines that improve their own prompts, a capability that amplifies both the power and the risk of the architecture.[3,12] The self-improving pipeline was therefore reported as the point where the architecture leaves the category of a fixed tool and enters the category of a changing workforce, because a tool that rewrites its own instructions is a tool whose behaviour the founder can no longer fully specify in advance.

In turn, the coordination of agents in dynamic environments was found to require adaptive role allocation rather than a fixed assignment, because the environment changes the task faster than a static role can follow it, and the token efficiency of that allocation was improved by a dynamic role-allocation mediator.[13,14] The same architectural principles were observed in adjacent literatures, in multi-robot coordination and in logistics, where the division of labour among agents followed the same logic of roles and messages, and the multi-robot survey confirmed that task allocation, not the individual capability, is the binding constraint of the system.(15-17) The architecture was also found to admit a state-aware coordinator for program repair, a mesh protocol for multi-context ecosystems, and a cognitive orchestration for patent intelligence, which together show that the division of labour scales beyond a single domain and into the knowledge work of the firm.(18-20) The smallest of the secondary results was that tool sharing among agents, the concurrent access to the same instruments, was the least standardized aspect of the architecture, with consistency and access coordination still open, and it was this, not the coordination schemes, that remained unresolved at the close of the review (Figure 1; Table 1).

The coordination schemes themselves were found to carry a measurable trade-off. Debate among agents improved the accuracy of a joint answer at the cost of more tokens and more time, while a single orchestrator with a fixed plan was cheaper but less robust to a surprise in the task.[6,14] The choice between the two was reported as a function of the task rather than as a property of the system, with the dynamic allocation of roles closing part of the gap between the two extremes.(13) The lifecycle survey framed the same trade-off as a workflow question, because the graph that connects the agents, not the model that each agent runs, determines how much the system spends to reach an answer.(4) The mesh protocol reported a further complication, because the agents in a complex ecosystem must also agree on the context they share, and the context, rather than the task, was found to be the source of the most persistent coordination failures.(19) The state-aware coordinator reported that the program-repair domain tolerates a fixed plan well, because the goal is known in advance, and this is the domain where the multi-agent system was closest to a drop-in workforce.(18) The patent-intelligence orchestration reported the opposite, because the goal there is open and the orchestration must adapt to the document rather than to a plan, and the two domains together mark the ends of the coordination spectrum.(20) The spectrum, not the average, is the datum that matters for the one-person company, because the founder's tasks span both ends. The consensus scheme added a further datum, because consensus among agents raised the reliability of a decision at the cost of requiring the agents to share a common representation, and the absence of that shared representation was reported as the point where the consensus scheme breaks, so the scheme is robust only where the agents already agree on the world.(6)

https://doi.org/10.62486/978-9915-9966-3-9.Ch11

Table 1. Architectural elements of multi-agent systems and their standardization.

Architectural element	Reported status	Representative studies
Role assignment and memory	Standardized	(1,4)
Message passing and coordination	Multiple schemes on shared primitives	(2,6)
Personas and self-improvement	Emerging, amplifies risk	(3,12)
Dynamic role allocation	Required in dynamic settings	(13,14)
Tool sharing and access	Least standardized	(19,21)

Note: Coordination schemes rest on the shared primitives of roles, memory and messages.

BUSINESS FUNCTIONS PERFORMED BY AGENTS

The second result concerns the business functions, and the main datum is that multi-agent systems now perform the bounded functions that constitute a firm's back office, while the open-ended judgment functions remain with the founder. It was observed in the enterprise literature that agentic business process management had become a research program in its own right, with a manifesto that treats the agent as the new unit of process execution, and the bibliometric evidence mapped the research on agentic artificial intelligence in small and medium enterprises and found a pattern of rapid growth concentrated in process automation.[7,8] The functions that were recorded as performed included marketing decision-making with auditable steps, vendor evaluation with risk-aware scoring, and workflow automation in cloud and enterprise resource planning environments, each reported with the caveat that the workflow was bounded and auditable.(22-24) The comparison of agentic automation against robotic process automation was found to favour the agent for tasks that require language understanding and to favour robotic process automation for deterministic tasks, a result that locates the agent's advantage precisely in the semi-structured middle of the back office.(25)

Similarly, a five-pillar framework for conversational enterprise management reported that the agent can coordinate across finance, human resources and procurement, and the integration of agentic systems into micro, small and medium enterprises was reported with the recurring constraint that the deployment is gated by the process, not by the model.[26,27] A human-centred entrepreneurship framework extended the same logic to the founding process of the solopreneur, and the coordination of agents for manufacturing task allocation under constraints was found to be a solved subproblem, with the constraint handled by the orchestration rather than by the model.[28,29] The automation of the labour market and of the back office was reported as an emerging application, with the agentic system described as a workforce for dynamic labour markets rather than as a static tool, and the integration of agentic systems into legacy enterprise systems was reported to require a mediation layer.[30,31] The human-in-the-loop evidence added that the process automation delivered by agents still requires a human to approve the exceptions, and the data-quality evidence added that the agent's output is only as good as the data it acts on.[32,33] The smallest of the secondary results was that the industry applications were reported as frameworks in greater number than as measured outcomes, with the study of agentic artificial intelligence in industry finding more architectures than evaluations (Table 2).

The evidence on the functions is uneven across the back office, and the unevenness is itself a finding. The procurement and marketing functions were reported with the strongest audit trails, because the workflow there is a sequence of comparisons that a model performs well and a human can check.[22,23] The finance and human-resources functions were reported at the framework level, because the exception handling and the compliance judgment in those functions resist a fixed workflow.(26) The process-automation function was reported as the most mature, because it is the closest to the robotic process automation baseline against which it is benchmarked, and the benchmark shows that the agent's advantage is real but bounded by the determinism of the task.[25,27] The founding function was reported as the least evidenced, because the human-centred entrepreneurship framework is a design and not yet a measured outcome, and the study of agentic artificial intelligence in industry confirmed that the founding and strategic judgments are where the architectures outnumber the evaluations.[28,34] The labour-market automation evidence added a caution, because a workforce of agents that is deployed across a dynamic labour market is a different object from a back office that automates a fixed process, and the former is where the oversight question becomes acute.(30) The pattern across the functions is therefore a gradient from the deterministic to the judgment-laden, and the one-

https://doi.org/10.62486/978-9915-9966-3-9.Ch11

person company that leans on the gradient inherits its steepest part. The gradient is the practical reading of the evidence, because a founder who automates the deterministic end of the back office gains the demonstrated benefit, and a founder who delegates the judgment-laden end loses the benefit that the evidence has not yet shown.(25,28)

Table 2. Business functions performed by multi-agent systems and the strength of evidence.

Business function	Evidence status	Representative studies
Process and workflow automation	Demonstrated, bounded	(8,24)
Marketing and procurement	Demonstrated with audit trails	(22,23)
Enterprise coordination	Framework-level, human-in-the-loop	(26,32)
Founding and strategic judgment	Least evidenced	(28,34)

Note: Bounded workflows carry the strongest evidence; open-ended judgment is least evidenced.

FAILURE MODES AND BOUNDARY CONDITIONS

The third result concerns the failure modes, and the main datum is that the multi-agent system inherits the failure modes of the single model and adds new ones that arise from coordination itself. It was observed that the attribution of a failure in a multi-agent system is itself an open problem, because the error of one agent propagates through the messages to the others, and the intention-behavior consistency of the system was used to automate that attribution only in part.[35] The collaboration was found to fail in a specific and new way, through the persuasion of one agent by another, and the adversarial influence that results was reported as a failure mode that a single model does not exhibit.[36] The security literature recorded seven cross-domain challenges, from prompt injection to agent identity, and a security-by-design framework for bounded agentic systems in enterprise environments was proposed in response, with the vulnerability-management literature reporting the same gap between the capability and the governance of the agent.[37-39]

In contrast, the hallucination problem was found to be reduced, though not removed, by multi-agent collaboration, and a conditional-mutual-information method was reported to detect collusion among agents in financial decision systems, a failure mode that is specific to the coordination rather than to the model.[40,41] The evaluation and regulation of agentic systems were found to lack a settled benchmark and a settled regulatory regime, with the critical assessment concluding that the capability is real and the oversight is lagging, and the workshop on trustworthy agentic systems reported the same gap from the engineering side.(9,10,42) The autonomy of the agent was distinguished from its agency by continuity, and bounded autonomy with verifiable safety was proposed as the design principle that reconciles autonomy with oversight.[43,44] The adaptive human oversight literature added that the balance between autonomy and oversight is itself a design variable rather than a fixed point, and the constraint-based analysis concluded that agentic artificial intelligence will not fully replace enterprise software in the near term precisely because the oversight and integration costs remain.[45,46] The smallest of the secondary results was that the workshop and the framework for trustworthy agentic systems were reported in greater number than the measured longitudinal safety of a deployed system (Table 3).

The boundary conditions follow from the failure modes, and they delimit the solopreneur's reliance in a specific way. The reliance is safe where the task is bounded, the output is auditable and the exception is caught by a human, and the reliance is unsafe where the task is open, the output is a judgment and the exception is silent.[44,45] The continuity distinction sharpens the bound, because an agent that acts with continuity across time is an agent whose errors accumulate, and the one-person company that runs such an agent is running a workforce whose mistakes compound.(43) The constraint-based analysis adds the economic bound, because the integration and oversight costs that the agent does not remove are the costs that a payroll used to hide, and the founder who replaces the payroll surfaces those costs onto a single person.(46)The security challenges add the external bound, because a multi-agent system that spans domains exposes the founder to a threat surface that a single-domain tool does not, and the vulnerability-management literature reports that the governance of that surface is itself lagging the capability.[37,39] The trustworthy-agentic workshop adds the engineering bound, because the property of trustworthiness is a design target that must be specified before it is measured, and the specification is missing from most of the frameworks.(42) The boundary conditions therefore

https://doi.org/10.62486/978-9915-9966-3-9.Ch11

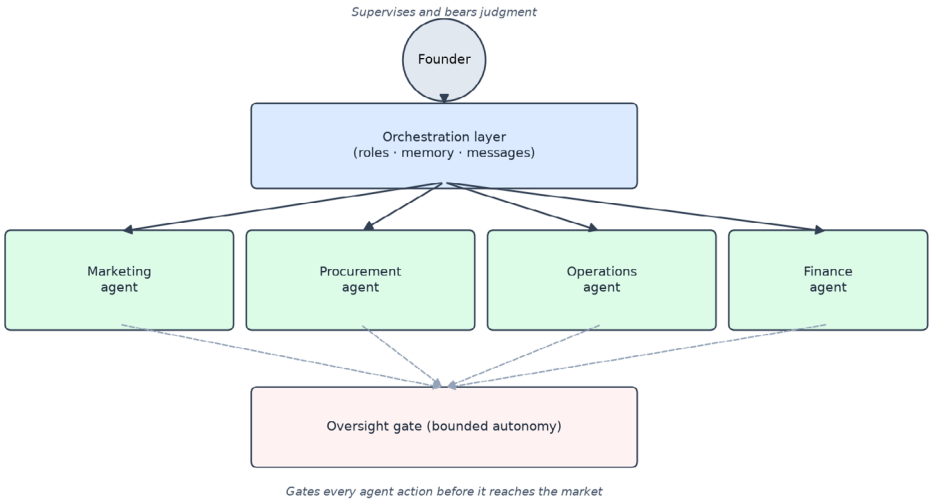
convert the failure modes into a decision rule, and the rule is that the founder may rely on the agent for the task it can be checked on and may not rely on it for the task it cannot. The auditability of the output is the operational test of the rule, because a task whose output can be checked is a task that can be delegated, and a task whose output is a judgment is a task that cannot, and the boundary between the two is the boundary that the founder must draw for every agent that the system adds.[44,45]

Table 3. Failure modes of multi-agent systems and their mitigation.

Failure mode	Mechanism	Mitigation reported
Error propagation	Message-passing amplifies one agent's error	Failure attribution(35)
Persuasion and collusion	Adversarial influence among agents	Detection methods(36,41)
Security gaps	Prompt injection, identity	Security-by-design(37,38)
Hallucination	Model-level, reduced not removed	Collaboration and verification(40)
Missing oversight	Benchmarks and regulation lag	Bounded autonomy and human oversight(9,45)

Note: Coordination-specific failure modes are added to the single-model failure modes.

Figure 1. Orchestration and division of labour in a multi-agent system supporting a one-person company.



Note: A single founder supervises specialized agents that coordinate through a shared orchestration layer and are gated by oversight.

DISCUSSION

The main finding of this review is that the multi-agent system is a real and rapidly maturing technical substrate for the workforce of the one-person company, and that its reliability is demonstrated for bounded business functions but not for the open-ended judgment that a founder cannot outsource. It was observed that the architecture has standardized on a division of labour among specialized agents, that the business functions now performed cover the back office of a small firm, and that the failure modes, error propagation, persuasion, collusion, security gaps and a missing oversight regime, bound the reliance that a solopreneur can place on the system. (1,8,9) The finding is therefore conditional, that the workforce exists as a technical substrate and does not yet exist as a substitute for judgment. The three results are ordered by their strength, with the architecture the most settled, the business function the most demonstrated and the failure mode the least governed, and the order of the three is itself the review's answer to the chapter's question, because a workforce is defined by the third result as much as by the first two.

The finding is concordant with the studies that report the standardization of the architecture and the breadth of the business functions. The survey literature agrees that the multi-agent

https://doi.org/10.62486/978-9915-9966-3-9.Ch11

system has moved from a single model to an orchestrated collection of roles, and the enterprise literature agrees that the collection performs the bounded functions of a firm, with the same division between the bounded and the unbounded appearing in both.(1,7,8) The benchmark comparison against robotic process automation agrees that the agent's advantage is language understanding, which is the middle of the back office, and the human-in-the-loop studies agree that the human remains the bearer of the exception.[25,32] The multi-robot and logistics literatures agree on the architectural claim, because the task allocation, not the individual capability, is the binding constraint in every domain where agents coordinate.[15,16] The concordance is strongest on the capability and weakest on the oversight, which is the point where the discrepant studies enter. The concordance extends to the detail of the architecture, because the surveys, the multi-robot literature and the logistics literature each report that the binding constraint is the allocation of tasks among agents rather than the capability of any single agent, and this is the same constraint that the business literature reports as the integration cost of the agentic system.[1,15,16] The concordance also extends to the division between the bounded and the unbounded, because the enterprise frameworks and the human-in-the-loop studies both place the exception, and not the routine, with the human, and the two placements are the same boundary drawn twice.[26,32] The standardization evidence is concordant with the same boundary, because the roles, memory and messages that the field standardized are the primitives of the bounded task, and the unbounded judgment has no primitive, which is why the standardization stops where the judgment begins.(1,5)

In contrast, the discrepant studies bound the finding on the margin where the solopreneur's reliance would be greatest. The constraint-based analysis reported that agentic artificial intelligence will not fully replace enterprise software in the near term, and the divergence is explained by the outcome considered, because the measured cost is the integration and oversight cost rather than the task capability, and a founder who replaces a payroll inherits that cost in full.(46) The benchmark evidence reported that the agent is not uniformly better than robotic process automation, and the divergence is explained by the task, because deterministic tasks do not reward the agent's language advantage.[25] The failure-mode evidence reported that collaboration itself introduces new failures, and the divergence is explained by the unit, because the single model's errors are compounded by the coordination that the multi-agent system adds.[36,37] Each discrepancy is a boundary condition in the terms of this review, and the boundary conditions together locate the gap between the substrate and the workforce. The collusion evidence is the clearest instance of a discrepancy that is new to the multi-agent setting, because a single model cannot collude with itself, and the fact that the multi-agent system introduces this failure means that the founder inherits a risk that did not exist in the single-model tool.(41) The persuasion evidence is the same point from a different direction, because the failure is not a bad output but a bad coordination, and the founder who audits the output rather than the coordination will miss it.(36) The two new failure modes are therefore not refinements of the old ones but additions, and they are the reason the discrepant studies cannot be dismissed as pessimism. The enterprise software point is the same bound stated in economic terms, because the software that the agent was supposed to replace is the software that still carries the integration and the audit, and the one-person company that hopes to skip that software inherits its cost in a less standardized form.(46)

The mechanisms that explain the finding are two, one technical and one organizational. The technical mechanism is the division of labour itself, because the decomposition of a goal into roles and messages is what lets the system scale, and the same decomposition is what lets an error propagate, so the architecture and the failure mode share a single mechanism.[1,35] The organizational mechanism is the shift of the exception to the founder, because the bounded workflow that the agent performs is bounded precisely by the judgment that it defers, and the founder who adopts the system inherits the sum of the deferred judgments.[32,45] The two mechanisms together explain why the workforce is real at the level of the task and absent at the level of the judgment. A third mechanism, the standardization mechanism, explains the rapid maturation of the substrate, because the convergence of the field on roles, memory and messages lowered the cost of comparing and reusing architectures, and the same standardization that accelerated the field is what makes the architecture a plausible foundation for a workforce rather than a research artefact.[1,5] The three mechanisms, division, exception and standardization, are therefore not alternative explanations but the three parts of a single account, in which the system scales because it divides, defers because it is bounded, and matures because it standardizes. The three mechanisms also explain the timing of the finding, because the field moved from a single model to a multi-agent system only when the primitives became standard enough to reuse, and

https://doi.org/10.62486/978-9915-9966-3-9.Ch11

the one-person company became thinkable only when the primitives became standard enough to trust, so the technical and the entrepreneurial claims share a single clock.

The review has limitations that bound the confidence of its conclusion, and each limitation carries the recommendation that would overcome it. The synthesis is narrative rather than systematic, so relevant work may have been missed, and the recommendation is a systematic search with a registered protocol once the construct is indexed consistently. The evidence is concentrated in the years 2024 to 2026, so the longitudinal safety and the survival of agent-supported ventures are unmeasured, and the recommendation is a longitudinal design that follows one-person companies over time. The business evidence is reported as frameworks in greater number than as measured outcomes, and the recommendation is a benchmark that measures the business function against a human baseline rather than against a prior framework. (34) The oversight evidence is conceptual and benchmark-based rather than operational, and the recommendation is an evaluation that follows a deployed system through a real failure.(9) Each limitation is a gap in the evidence rather than a gap in the argument, and the conclusion is offered as the best current assembly. A further limitation is that the review did not separate the benchmark-based evidence from the deployment-based evidence with a formal quality appraisal, and the recommendation is a risk-of-bias assessment adapted to the framework and benchmark designs that dominate the field. The review also did not search the grey literature of vendor reports, and the recommendation is a structured search of that literature once a protocol for its quality exists, because the frontier of deployment may currently be reported there rather than in peer-reviewed venues. The heterogeneity of the field is a further limitation, because the studies span benchmarks, frameworks and deployments with no common unit, and the recommendation is a common reporting standard for the business function, the failure rate and the oversight cost of an agentic system, so that the next synthesis can combine rather than merely juxtapose the studies.

The implications are practical for the founder and the regulator. The founder who adopts a multi-agent workforce should treat it as a bounded back office with a human in the loop, not as a substitute for the judgment, the liability and the oversight that the firm still requires, and the regulator who governs the one-person company should require the oversight that the technology defers. The recommendation that follows from each limitation is therefore the same, to measure what the framework promises against a baseline that a human sets, and to treat the workforce as a substrate that the founder must supervise rather than a substitute that the founder can trust. The research agenda that follows is specific, to benchmark the business functions against a human baseline, to follow a deployed multi-agent system through a real failure, and to measure the survival of the one-person companies that adopt the workforce, because the three measures together would convert the review's conditional finding into an estimate. The one-person company, in the meantime, is best described as a founder who supervises a bounded back office of agents rather than as a founder who has been replaced by them. The founder's oversight obligation is therefore not a temporary cost that the technology will soon remove but a permanent feature of the architecture, because the division of labour that gives the system its power is the same division that leaves the founder with the judgment, the exception and the liability, and a workforce that is real at the task and absent at the judgment is a workforce that a founder supervises rather than a workforce that a founder is. The distinction between supervising and being is the practical core of the chapter, because the one-person company that survives is the one that keeps the supervision, and the one that fails is the one that confuses the substrate for the substitute. That distinction is the condition under which the workforce claim holds, and it is the claim the review set out to delimit.

CONCLUSIONS

Multi-agent artificial intelligence systems have standardized on a division of labour among specialized agents and have thereby become a plausible technical substrate for the workforce of the one-person company. The business functions that these systems perform are demonstrated for bounded workflows, and the open-ended judgment that a founder cannot outsource remains the founder's. The solopreneur who adopts a multi-agent workforce therefore inherits a set of oversight, security and liability obligations that the technology does not discharge.

https://doi.org/10.62486/978-9915-9966-3-9.Ch11

REFERENCES

Bajoria S, Ranjan S, M A, Suri V, Hassija V, Chamola V, et al. From Language Models to Agentic AI: A Survey of Autonomous, Action-Enabled, and Collaborative LLM Agents. CognComput. 2026;18(1):103. doi: 10.1007/s12559-026-10619-1
Cheng Y, Zhang C, Zhang Z, Meng X, Hong S, Li W, et al. Exploring Large Language Model-Based Intelligent Agents: Definitions, Methods, and Prospects. WIREs Data Min & Knowl. 2026;16(3):e70111. doi: 10.1002/widm.70111
İbrahimoğlu N, Demirezen MU, Yıldız F. Autonomous AI agents and recursive self-improvement: a comprehensive survey of contemporary architectures, capabilities, and safety frameworks. Cluster Comput. 2026;29(13):766. doi: 10.1007/s10586-026-06537-4
Kang C, Li K. A lifecycle-oriented survey of LLM-based agents: Workflow, graph augmentation, and future directions. Neurocomputing. 2026;704:134838. doi: 10.1016/j.neucom.2026.134838
Alva LR, Pandey B. Agentic AI systems in the age of generative models: architectures, cloud scalability, and real-world applications. Artif Intell Rev. 2026;59(3):88. doi: 10.1007/s10462-025-11458-6
Hossain E, Nipu MMHB, Mahmood MS, Hossen MJ, Mridha MF. Safe and Scalable Collaboration in Multiagent LLM Systems: A Comprehensive Review. IEEE Trans Syst Man Cybern, Syst. 2026;56(10):6139-6155. doi: 10.1109/tsmc.2026.3704902
Olujimi PA, Owolawi PA, Pretorius A, Van Wyk E. Mapping the research landscape of agentic AI in SMMEs through a bibliometric analysis of patterns and knowledge gaps. DiscovArtif Intell. 2025;6(1):63. doi: 10.1007/s44163-025-00764-1
Calvanese D, Casciani A, De Giacomo G, Dumas M, Fournier F, Kampik T, et al. Agentic Business Process Management: A research manifesto. Information Systems. 2026;140:102738. doi: 10.1016/j.is.2026.102738
Farooq A, Raza S, Karim N, Iqbal H, Vasilakos AV, Emmanouilidis C. Evaluating and regulating agentic AI: A study of benchmarks, metrics, and regulation. Information Fusion. 2026;136:104444. doi: 10.1016/j.inffus.2026.104444
Khan RU, Aljuaid H, Ning Z. A critical assessment of the rise of agentic AI, its capabilities, task domains and open challenges. DiscovArtif Intell. 2026;6(1):1343. doi: 10.1007/s44163-026-02342-5
Dritsas E, Trigka M. Intelligent agents: architectures, learning paradigms, deployment models, and applications. Evol Intel. 2026;19(5):145. doi: 10.1007/s12065-026-01259-3
Nawaz AJ, Patock JR, Valencia I, Li B, Berkowitz JS, Coxe S, et al. A framework for cognitive personality priors in agentic AI: the role of personas in autonomous team science. BioData Mining. 2026;19(1):75. doi: 10.1186/s13040-026-00598-2
Wu X, Wang Z, Zhang J, Gao Z, Jiang Z. Semantic-Driven Adaptive Multi-Agent Collaboration via LLMs in Dynamic Environments. 2026 International Conference on Generative Artificial Intelligence and Information Security (GAIIS). 2026:119-123. doi: 10.1109/gaiis69281.2026.11519182
Chen Z, Zhen H, Zhu Z, Jiang J, Kong M, Gu X. D-RAM: A Dynamic Role-Allocation Mediator for Improving Token Efficiency in Multi-Agent Collaborative Software Development. 2026 IEEE 8th International Conference on Communications, Information System and Computer Engineering (CISCE). 2026:1350-1353. doi: 10.1109/cisce69494.2026.11504945
Li P, An Z, Abrar S, Zhou L. Large language models for multi-robot systems: a survey. Auton Robot. 2026;50(3):30. doi: 10.1007/s10514-026-10257-4
Sorooshian S, Ahadi N, Liravi P. From multi-agent logistics to agentic artificial intelligence: review of autonomous coordination in Q-commerce-related last-mile fulfilment. Electronic Commerce Research and Applications. 2026;79:101631. doi: 10.1016/j.elerap.2026.101631
Bakhat SID, Ullah I, Bai X, Zhang W, Li J, Ge SS. Advancing Multirobot Task Allocation: A Comprehensive Review of Emerging Trends and Challenges. IEEE Internet Things J. 2026;13(19):44021-44058. doi: 10.1109/jiot.2026.3699696
Wang S, Lu L, Qiu S. SCAgent : An LLM -Based Agent With State-Aware Coordination for Automated Program Repair. Expert Systems. 2026;43(10):e70400. doi: 10.1111/exsy.70400
Ginanjar A, Li X. Mesh Multi-Context Protocol: An Adaptive Agentic AI With Mesh Concept for A Complex Ecosystem. Applied AI Letters. 2026;7(3):e70041. doi: 10.1002/ail2.70041
Wang HY, Ke HR. Agentic cognitive orchestration for cross-lingual patent intelligence: A modular workflow automation approach. World Patent Information. 2026;86:102483. doi: 10.1016/j.wpi.2026.102483
Xiang Y. Multi-Agent Tool Sharing: Task Allocation, Concurrent Access Coordination, and Consistency Guarantees in Collaborative LLM Agent Systems. 2026 International Conference on Large Models and Intelligent Agent Technologies (LMIAT). 2026:85-89. doi: 10.1109/lmiat70705.2026.11682229
Liu Q, Cui H, Wang Y, Lee CS, Chen S, Yang L. A Multi-Agent Large Language Model Framework for Marketing Decision-Making with Auditable Attribution Analysis. Proceedings of the 2026 International Conference on Artificial Intelligence and Control. 2026:606-612. doi: 10.1145/3807246.3807341
Joshi A, Singh M, Parashar D, Singh M. A multi-agent large language model framework for intelligent vendor evaluation and risk-aware procurement decisions. Sci Rep. 2026;16(1):19826. doi: 10.1038/s41598-026-50952-x
Gundala M, Sura NK, Hakeem M. An Agentic AI Framework for Enterprise Workflow Automation in Cloud Environments. 2026 International Conference on Intelligent and Sustainable AI Systems (ICOSAAS). 2026:874-880. doi: 10.1109/icosaas68663.2026.11649277
Průcha P, Matoušková M, Strnad J. Are LLM agents the new RPA? A comparative benchmark study of agentic automation and RPA across standardized workflow tasks. Cogent Business & Management.

2026;13(1):2724760. doi: 10.1080/23311975.2026.2724760
Assidiqi MH, Alghazzawi D, Alarifi S, Cheng L. Conversational Enterprise Management Through Agentic AI: A Five-Pillar Framework for LLM-Driven ERP Systems. Applied Artificial Intelligence. 2026;40(1):2700868. doi: 10.1080/08839514.2026.2700868
Parthasarathy VP. Agentic AI Integration for Process Automation in MSMEs. 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC). 2026:1-6. doi: 10.1109/icaic67076.2026.11395804
Gao Y, Piao Y, Xuan D. HCEA: A Multi-Agent Framework for Sustainable Human-Centered Entrepreneurship Based on a Large Language Model. Sustainability. 2026;18(7):3554. doi: 10.3390/su18073554
Ye W, Li X. Coordination of LLM-Embodied Agents for manufacturing task allocation under constraints. Journal of Manufacturing Systems. 2026;87:17-26. doi: 10.1016/j.jmsy.2026.04.028
Sharma S, Sharma S. Agentic Artificial Intelligence-Powered Automation for Dynamic Labor Markets and Business. In: Advances in Computational Intelligence and Robotics. IGI Global Scientific Publishing; 2026. p. 153-186. doi: 10.4018/979-8-3373-5816-1.ch006
Mohanam BN. BridgeAgent: A Graph-Aware Mediation Framework for Connecting Agentic AI to Legacy Enterprise Systems via Adaptive API Translation. 2026 International Conference on Integrated Intelligence and Cognitive Engineering (ICIICE). 2026:1-6. doi: 10.1109/iciice69672.2026.11565428
Alla PB. Augmenting Intelligent Process Automation through Generative AI for Human-in-the-Loop Decision Systems. Digital Engineering. 2026;8:100071. doi: 10.1016/j.dte.2025.100071
Cravero A, Díaz-Villegas J, Xiao Z. From Data Quality to Quality of Agentic Data Use: A Conceptual Framework for Agentic Data Engineering. Appl Sci. 2026;16(17):8887. doi: 10.3390/app16178887
Kukreja S, Bharate V, Kamble S. A Study of Agentic AI in Industry Applications. 2026 IEEE International Conference on AI Engineering and Innovations (AIEI). 2026:1-8. doi: 10.1109/aiei69164.2026.11496688
Wu H, Zheng W, Zhang M, Bai X, Pu M, Sun L, et al. Intention-behavior consistency-based automated failure attribution for LLM-driven multi-agent systems. Expert Systems with Applications. 2026;331:133044. doi: 10.1016/j.eswa.2026.133044
Kraidia I, Qaddara I, Almutairi A, Alzaben N, Belhouari SB. When collaboration fails: persuasion driven adversarial influence in multi agent large language model debate. Sci Rep. 2026;16(1):11640. doi: 10.1038/s41598-026-42705-7
Ko R, Chung J, Zheng S, Xiao C, Kim TW, Onizuka M, et al. Seven security challenges in cross-domain multi-agent LLM systems. npjArtif Intell. 2026;2(1):72. doi: 10.1038/s44387-026-00128-9
Paul S. A Security-by-Design Framework for Bounded Agentic AI in ERP-Integrated Smart Semiconductor and Electronics Manufacturing. 2026 6th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET). 2026:1-6. doi: 10.1109/iraset68627.2026.11538696
Akrami Y, Malkawi M, Alhajj R. Large Language Models and Agentic AI for Vulnerability Management: A Systematic Review. Expert Systems. 2026;43(11):e70437. doi: 10.1111/exsy.70437
Shen S, Sun D, Qin Z, Su Y, Zhu L, Lu J, et al. Towards higher quality and fewer hallucinations: A multi-agent collaboration framework for LLMs. Information Processing & Management. 2026;63(8):104953. doi: 10.1016/j.ipm.2026.104953
Nait Cherif A, Youssfi M, Bouattane O. Detecting Collusion in LLM-based Multi-agent Financial Decision Systems: A Conditional-mutual-information Detector and a Modular Oversight Architecture. IJIES. 2026;19(9):1379-1391. doi: 10.22266/ijies2026.0930.66
Dinakar K, Weisz JD, Lieberman H, Geyer W. AgentCraft: Workshop on Developing Trustworthy Agentic AI Systems. Companion Proceedings of the 31st International Conference on Intelligent User Interfaces. 2026:240-242. doi: 10.1145/3742414.3794957
Kahl P. How continuity distinguishes autonomy from agency in agentic AI. DiscovArtif Intell. 2026;6(1):1187. doi: 10.1007/s44163-026-01675-5
Ramaswamy S, Nayak D. Bounded Autonomy and Verifiable Safety for Agentic AI Enabled Automation. J Intell Robot Syst. 2026;112(4):93. doi: 10.1007/s10846-026-02467-w
Kumar M, Singh D. Balancing autonomy and oversight in reliable agentic artificial intelligence through adaptive human interaction architectures. DiscovArtif Intell. 2026;6(1):709. doi: 10.1007/s44163-026-01373-2
Zaa J. Why Agentic AI Will Not Fully Replace Enterprise SaaS in the Near Term: A Constraint-Based Feasibility Model. 2026 IEEE International Conference on Electro Information Technology (eIT). 2026:091-096. doi: 10.1109/eit68936.2026.11670371