The Rise of the Zero-Human Company: Can AI run a company without employees?

Christian Grant
Deep AI Mind

March 15, 2026

The Rise of the Zero-Human Company: Can AI run a company without employees?

Executive summary

- A “zero-human company” is best understood as an **operating model** in which **AI agents execute most day-to-day work** (research, production, sales ops, and support) while **humans—if present at all—sit outside the workflow as overseers** (owners, directors, auditors) rather than employees. The most credible early evidence comes from structured experiments (e.g., a Big-4–adjacent research initiative) rather than social-media claims. [1]

- The concept is emerging now because the **unit cost of “cognitive labor” is dropping** and agent platforms increasingly combine (1) **tool access**, (2) **workflow structure**, and (3) **multi-agent coordination**, which together make end-to-end automation more plausible than “chatbot-era” automation. [2]

- The technological inflection is not “LLMs got smarter” alone; it is that ecosystems are converging on **standardized integration and interoperability**: protocols to connect agents to tools/context (e.g., MCP) and to connect **agents to other agents** across vendors (e.g., A2A). [3]

- The biggest near-term opportunity is **execution compression**: faster iteration cycles, lower overhead, and new “tiny-team” or solo-founder structures where agents absorb workflow toil (drafting, scheduling, reporting, outreach, triage). Evidence includes survey-reported increases in AI and agent adoption and observational accounts of founders using multi-agent “councils.” [4]

- The biggest near-term constraint is **organizational reliability**: when agents are assigned broad human job titles one-to-one (e.g., “CFO agent”), they may drift, hallucinate, or stall—forcing a redesign toward **process-mapped micro-tasks** and stronger controls. [5]

- The most acute systemic risk is **security**—especially when agents gain broad file/network permissions and install third-party “skills” from open registries. Real incidents show malware distribution via agent skill marketplaces, and platform documentation itself warns to treat third-party skills as untrusted. [6]

- Economically, zero-human firms could increase **venture formation** (more “shots on goal”) and intensify competition for scarce attention and distribution. Venture

funding and “newly funded AI companies” metrics are already large and rising in key regions, but this does not guarantee durable businesses. [7]

- Bear in mind that the path from the DARPA Grand Challenge in 2004 to meaningful autonomous vehicle deployment in major U.S. cities took nearly two decades, not a sudden leap. DARPA’s first Grand Challenge took place in March 2004; Waymo began welcoming riders in San Francisco in 2021 and started removing the human driver there in 2022. The likely trajectory over the next 10–20 years is therefore not a clean jump to fully autonomous corporations, but a spectrum: (1) AI-assisted companies (tools), (2) hybrid human-AI firms (agents + people), and (3) mostly autonomous operator models in bounded domains with strong controls, auditability, and legal accountability. [8]

Concept definition

A **Zero-Human Company** is an organization whose **core operating workflows**—from sensing opportunities to delivering products and handling customers—are **executed by AI agents** rather than employees. “Zero-human” is therefore more precise as an **internal labor statement** (who performs work) than a legal one (who owns/liably controls the entity). The cleanest operational definition is: **humans may exist as owners/supervisors**, but they are not required for routine execution loops. [1]

How it differs from traditional firms

Traditional firms coordinate work through a mix of **human roles**, software systems, and managerial processes. In “zero-human” designs, the coordination substrate shifts: agents (and their orchestration layer) become the **primary executors** while “management” becomes increasingly **algorithmic**—a set of rules, policies, and monitoring mechanisms used to instruct, evaluate, or constrain work. This is conceptually adjacent to “algorithmic management,” which the OECD documents as already widespread for automating or supporting managerial tasks, and which raises questions of accountability and transparency. [9]

Relationship to agents and automation

- **AI tools** augment humans: copilots, summarizers, code assistants, analytics helpers. [10]
- **AI-run organizations** move beyond augmentation: they depend on agents that can **plan, execute multi-step workflows, call tools, recover from errors**, and coordinate with other agents through structured patterns and protocols. [2]

In practice, early implementations often land between these poles: companies report widespread AI usage while still struggling to scale to enterprise-wide impact; agent scaling remains limited in most functions, indicating the organizational leap is non-trivial. [11]

Startup model implications in one table

Dimension	Traditional startup	Agent-native "zero-human" trajectory
Primary constraint	Hiring, coordination, cash runway	**Distribution & trust**; **controls & security**; model/tool costs [12]
Execution loop	Humans plan → build → ship	Agents "work the pipeline" via chained workflows + tool calls [13]
Failure mode	Team bottlenecks, slow iteration	Agent drift/hallucination; compounding errors; unsafe tool use [14]
Governance	People-managed processes	Policy engine + audits + human supervisors outside workflow [15]

Technology drivers and the evolution of multi-agent platforms

The phrase "AI runs the company" can be misleading: what matters is the **stack**—LLMs + orchestration + tools + security + interoperability + marketplaces. Recent platform directions show convergence on a few primitives.

LLMs as "augmented" decision engines

A practical definition used by leading builders is that agents are **LLMs using tools in a loop**; the "augmented LLM" combines retrieval, tools, and memory, enabling models to generate search queries, select tools, and decide what to retain. [16]

Crucially, the move to "company-level autonomy" depends less on eloquent text and more on engineering for **tool reliability, context efficiency, and error recovery**—because business workflows involve long horizons, state, and messy data. [17]

Agent workflow patterns that scale beyond demos

A recurrent empirical lesson is that "unstructured autonomy" is fragile; reliable systems compose a handful of patterns:

- **Prompt chaining** (stepwise decomposition with checks), **routing** (dispatch to specialized paths), **parallelization** (sectioning/voting), **orchestrator-workers** (dynamic delegation), and **evaluator-optimizer** (critique loops). [18]
- Multi-agent architectures often instantiate these patterns explicitly: "subagents," "handoffs," "skills," "routers," and custom graphs for deterministic control. [19]

These patterns are not academic ornamentation: they are the difference between "an agent that chats" and a system that can repeatedly execute production work without compounding mistakes. [20]

Tooling, APIs, and cloud infrastructure: the "action surface"

Agents become economically meaningful when they act on real systems—CRMs, billing, deployment, support desks—through APIs and execution environments. Two emerging design directions are notable:

1. **Protocol standardization for tools/context (MCP).** MCP defines a client-host-server architecture to connect LLM applications to external tools and resources, aiming to reduce bespoke integrations. [21]
2. **Context and tool efficiency engineering.** "Programmatic tool calling" and "tool search" approaches aim to reduce context pollution and latency in multi-tool workflows, letting orchestration logic run in code and returning only final results to the model. [22]

Inter-agent communication and interoperability (A2A)

A core blocker to "AI-run organizations" is that real companies use **many systems**—and future agent stacks will likely be multi-vendor. An explicit push toward interoperability is the Agent2Agent protocol, presented as enabling agents built by different vendors/frameworks to communicate, exchange information securely, and coordinate actions across enterprise application estates. The protocol is described as built on existing standards (HTTP, SSE, JSON-RPC) and positioned as complementary to MCP (tools/context). [23]

From a "zero-human company" lens, A2A is less about chat and more about **organizational topology**: it is a candidate standard for how "departments" (sets of agents) coordinate across tool boundaries without humans relaying messages. [24]

Skills, marketplaces, and the composable organization (OpenClaw as an example)

A key platform evolution is the rise of **skill packaging**—portable bundles that "teach" agents how to use tools and workflows. For example, OpenClaw's documentation describes skills as directories containing `SKILL.md` with YAML frontmatter and instructions, loaded from bundled, local, or workspace locations, with explicit precedence rules. [25]

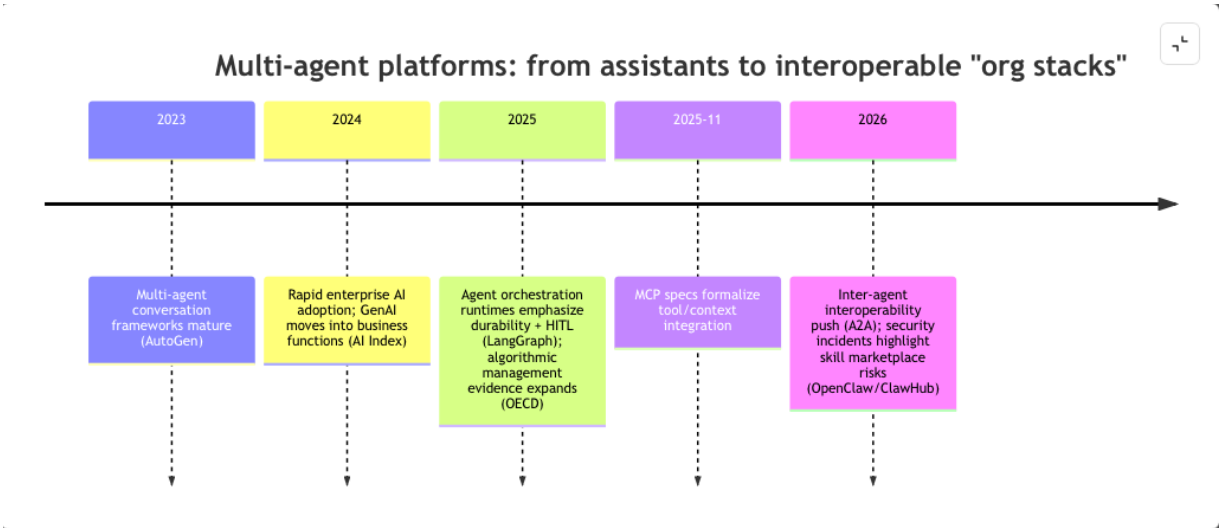
This model naturally produces **marketplaces/registries**, such as ClawHub, described as a public registry for skills with versioning, discovery, and moderation hooks. [26]

But the same mechanism creates a security fault line: platform documentation explicitly advises treating third-party skills as untrusted code, and recent reporting documents malware distributed through skill registries. [27]

A lightweight evolution timeline

```
Timeline title Multi-agent platforms: from assistants to interoperable "org stacks"
```

```
  2023 : Multi-agent conversation frameworks mature (AutoGen)
  2024 : Rapid enterprise AI adoption; GenAI moves into business functions (AI Index)
  2025 : Agent orchestration runtimes emphasize durability + HITL (LangGraph); algorithmic management evid
ence expands (OECD)
  2025-11 : MCP specs formalize tool/context integration
  2026 : Inter-agent interoperability push (A2A); security incidents highlight skill marketplace risks (Op
enClaw/ClawHub)
```



The sequence above reflects widely cited platform directions (multi-agent frameworks, orchestration runtimes, protocols) and the parallel emergence of governance/security constraints. [28]

Case studies and experiments

The KPMG Netherlands[29]–University of Amsterdam[30] zero-person experiment

This experiment is unusually valuable because it is framed as a **scientific study** (governance, liability, ethics) rather than a marketing demo. In the published description, the agent team selected the business concept (a webshop for personalized AI art), worked continuously, and produced business plans; the CEO agent's decision was accompanied by a **human supervisory board** that "monitors and coaches." [1]

The most important reported failure was organizational: when agents were mapped **one-to-one onto human roles** (e.g., CFO), they "drift off and hallucinate," ignore instructions, or stop working—forcing a redesign toward detailed process mapping and assigning agents to **small tasks** within processes. This is a concrete, real-world pointer that "AI workers" may need to be designed more like microservices than like employees with job titles. [5]

AI-assisted micro-companies: “councils,” not corporations

A recurring pattern in founder narratives is not literal “zero humans,” but **agent councils** that replace staff functions (legal, HR, finance) with specialized agents—often saving time but still requiring human judgment for high-stakes decisions. One example describes a solo founder using ~15 custom agents (including a “chief of staff”) to cover corporate functions, while emphasizing configuration, role authority, and cross-validation to reduce hallucinations. Treat this as **illustrative**, not representative. [31]

This pattern aligns with survey data: organizations widely report AI usage, many are experimenting with agents, but scaling remains limited and uneven across functions—consistent with “micro-company augmentation” happening faster than “fully autonomous firm operation.” [11]

Platforms and “company-in-a-box” narratives: value vs. attention

The most visible “zero human company” claims are emerging from podcasts and social channels rather than from audited disclosures. For example, a widely circulated episode summary describes experiments where agent-driven projects generate revenue and platforms launch many autonomous startups; it also highlights the core tension: lower execution cost may just create more competition for finite attention. These claims should be read as **directional signals** of experimentation, not verified performance data. [32]

The Reddit[33] discussion in r/accelerate points to similar skepticism: commenters suggest “it might already exist” but “doesn’t mean it’s going to make any money,” while estimating viability for some use cases on a short horizon. This is weak evidence, but it is useful as a market sentiment indicator: people intuit that automation may produce *entities*, not necessarily *value*. [34]

A cautionary platform case: skill marketplaces as attack surfaces

Open skill registries are a double-edged sword. ClawHub is explicitly designed as a skill registry with search and moderation hooks. [35]

Yet reporting documents malware distributed via OpenClaw skills, including instructions that lead users/agents to execute malicious commands, stealing secrets such as API credentials and SSH logins. OpenClaw’s own security guidance frames its model as a “personal assistant trust model,” not a hostile multi-tenant boundary—reinforcing that “agent companies” operating in adversarial environments need more robust security primitives than typical hobbyist stacks. [36]

Architecture of an autonomous company

A credible “AI-run company” design should be described less as a single super-agent and more as an **operating system**: orchestration + policies + tools + memory + evaluation + security.

Reference architecture: layers and responsibilities

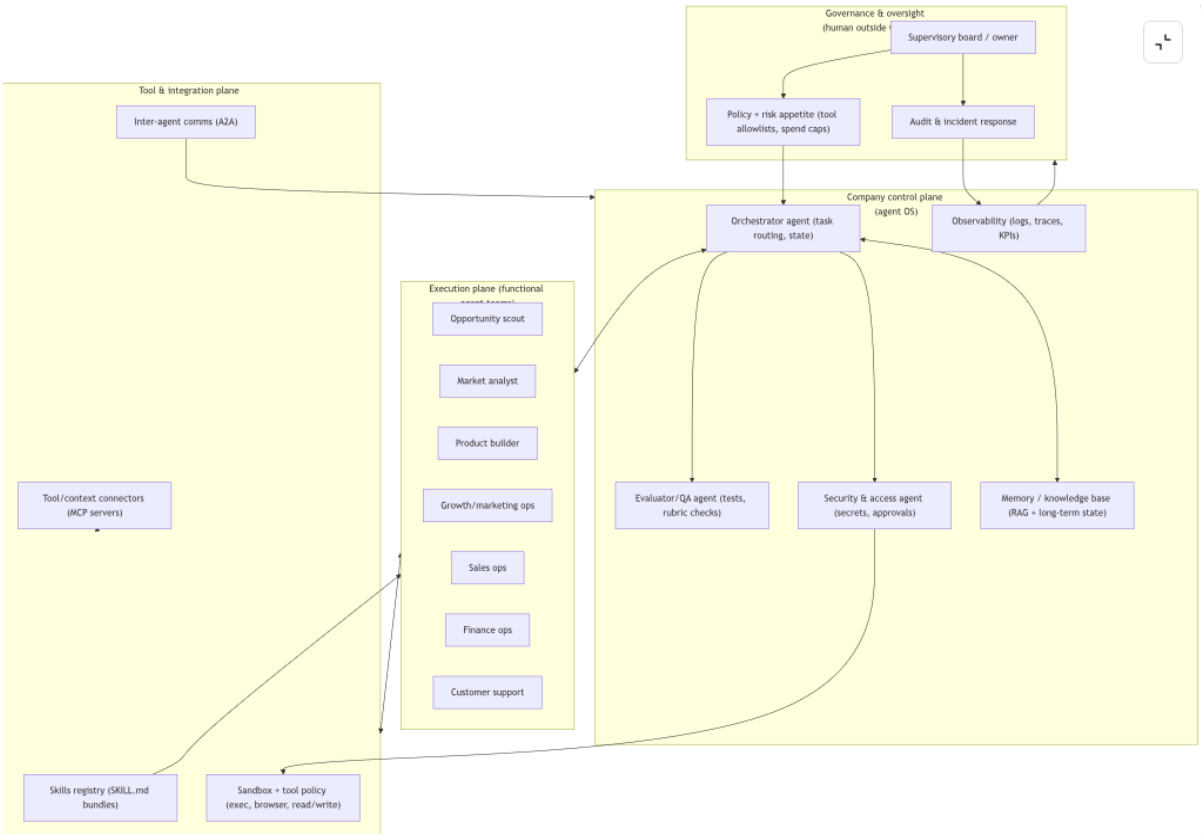
```
flowchart TB
  subgraph Gov["Governance & oversight (human outside workflow)"]
    BOD["Supervisory board / owner"]
    AUDIT["Audit & incident response"]
    POLICY["Policy + risk appetite (tool allowlists, spend caps)"]
  end

  subgraph Control["Company control plane (agent OS)"]
    ORCH["Orchestrator agent (task routing, state)"]
    EVAL["Evaluator/QA agent (tests, rubric checks)"]
    SEC["Security & access agent (secrets, approvals)"]
    OBS["Observability (logs, traces, KPIs)"]
    MEM["Memory / knowledge base (RAG + long-term state)"]
  end

  subgraph Work["Execution plane (functional agent teams)"]
    OPPO["Opportunity scout"]
    MKT["Market analyst"]
    PROD["Product builder"]
    GROW["Growth/marketing ops"]
    SALES["Sales ops"]
    FIN["Finance ops"]
    SUP["Customer support"]
  end

  subgraph Tools["Tool & integration plane"]
    SKILLS["Skills registry (SKILL.md bundles)"]
    MCP["Tool/context connectors (MCP servers)"]
    A2A["Inter-agent comms (A2A)"]
    SANDBOX["Sandbox + tool policy (exec, browser, read/write)"]
  end

  BOD --> POLICY --> ORCH
  BOD --> AUDIT --> OBS
  ORCH <--> Work
  ORCH --> EVAL
  ORCH --> SEC
  ORCH <--> MEM
  Work <--> Tools
  SEC --> SANDBOX
  SKILLS --> Work
  MCP --> Tools
  A2A --> Control
  OBS --> Gov
```



This blueprint is directly motivated by: (1) the need to decompose work into processes and micro-tasks (rather than job-title agents) to reduce drift; (2) the centrality of structured workflows (routing, chaining, evaluator loops); and (3) the reality that tool access and sandboxing define the true risk boundary. [37]

Explicit key agent roles (and why "micro-roles" win)

The KPMG/UvA experiment reports that mapping agents one-to-one to broad human roles (e.g., "CFO agent") led to hallucination and stalling; they shifted to **detailed process mapping** and assigning agents to small tasks. [1]

A practical compromise is to define *micro-roles* aligned to workflows and controls:

Agent role	Primary outputs	Typical skills/tools
Orchestrator ("chief of workflow")	Task decomposition, routing, state machine	Orchestrator-workers; routing graphs; durable execution [38]
Evaluator/QA	Rubric checks, regression tests, "stop/go" gates	Evaluator-optimizer loops; voting/parallel review [39]

Agent role	Primary outputs	Typical skills/tools
Security & access	Tool allow/deny, sandbox enforcement, secrets hygiene	Per-agent tool policies; sandboxing modes; prompt-injection guardrails [40]
Market intelligence	Competitor maps, pricing scans, customer insights	Tool-using loops + retrieval; controlled web tools [41]
Product builder	Specs → code/content → deploy	Programmatic tool calling; code execution with constrained context [42]
Growth ops	Campaign production, SEO/content pipelines	Prompt chaining; parallelization for variants; skills bundles [43]
Sales ops	Lead enrichment, outreach sequencing, CRM updates	Tool calling + approvals; strict messaging policies [44]
Finance ops	Invoicing, spend caps, reconciliation	Programmatic tool calling; budget compliance patterns [45]
Customer support	Triage → resolution → escalation	Routing by issue type; safe tools; human escalation gates [46]

Agent components that matter most in "company mode"

1. **State and durability.** Long-running agents need the ability to persist through failures and resume. LangGraph explicitly positions durable execution and human-in-the-loop state inspection as core benefits for stateful agents. [47]
2. **Skills/tool registry as the "app store of labor."** OpenClaw's model—skills with explicit format and gating rules plus a public registry—illustrates how companies could "install capabilities" rather than hire. It also illustrates the security costs of such extensibility. [48]
3. **Per-agent sandboxing and tool restriction.** OpenClaw's multi-agent tooling shows concrete mechanics: sandbox mode/scope (session/agent/shared), workspace access (none/ro/rw), and allow/deny lists (e.g., "read-only family agent"). This is a useful template for least-privilege design in autonomous firms. [49]
4. **Context efficiency and tool correctness.** Toolset bloat and ambiguous tool definitions are documented as common failure modes; advanced patterns like programmatic tool calling aim to keep intermediate data out of model context and reduce error-prone natural-language juggling. [50]

Communication between agents inside a company

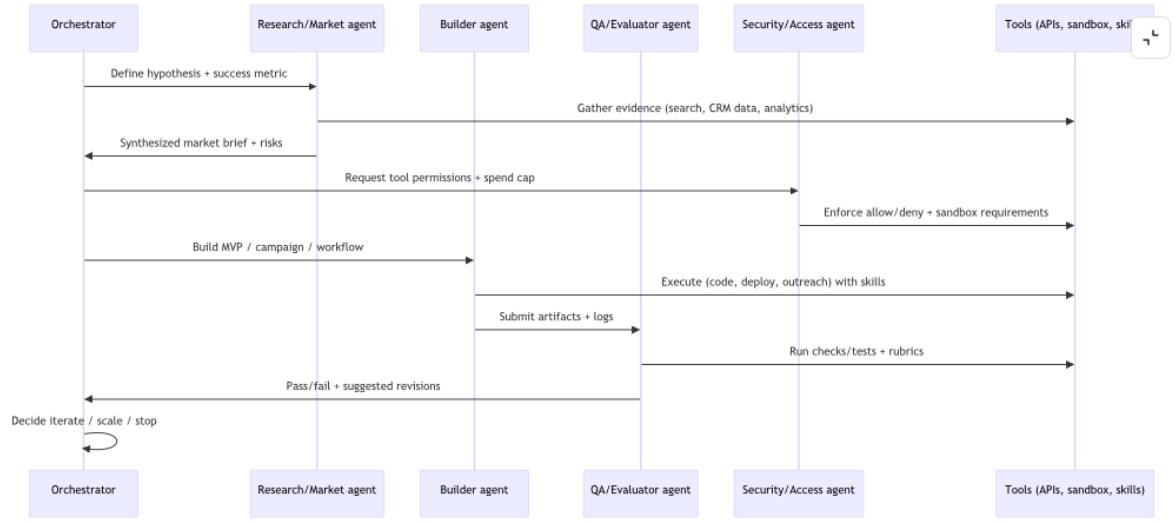
There are three practical "lanes" of agent communication:

- **Intra-runtime delegation (subagents).** OpenClaw subagents are described as background runs spawned from an existing agent run, executing in their own session IDs and "announcing" results back to the requester channel, with non-blocking spawn and delivery retries. This is a concrete blueprint for asynchronous internal delegation. [51]
- **Cross-system tool mediation (MCP).** MCP provides a standardized mechanism for agents to access tools and resources through a client-host-server architecture, helping avoid bespoke integrations in complex organizations. [21]
- **Cross-vendor agent interoperability (A2A).** A2A is explicitly framed as enabling agents built in different frameworks/vendors to communicate, exchange information, and coordinate actions, built on web standards and positioned as complementing MCP. In a "zero-human company," A2A is a candidate "department-to-department" protocol. [52]

A canonical workflow pattern for "company loops"

```
sequenceDiagram
  participant O as Orchestrator
  participant R as Research/Market agent
  participant B as Builder agent
  participant Q as QA/Evaluator agent
  participant S as Security/Access agent
  participant T as Tools (APIs, sandbox, skills)

  O->>R: Define hypothesis + success metric
  R->>T: Gather evidence (search, CRM data, analytics)
  R->>O: Synthesized market brief + risks
  O->>S: Request tool permissions + spend cap
  S->>T: Enforce allow/deny + sandbox requirements
  O->>B: Build MVP / campaign / workflow
  B->>T: Execute (code, deploy, outreach) with skills
  B->>Q: Submit artifacts + logs
  Q->>T: Run checks/tests + rubrics
  Q->>O: Pass/fail + suggested revisions
  O->>O: Decide iterate / scale / stop
```

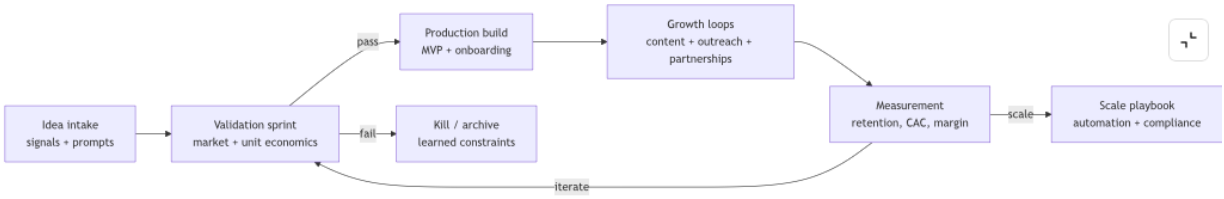


This sequence is an applied mix of: routing, orchestrator-workers delegation, evaluator-optimizer loops, and least-privilege tool execution—each documented as a practical pattern for reliable agent systems. [53]

"AI venture factory" blueprint (startup model integration)

A venture factory is the *organizational* version of the workflow above: a pipeline that produces many small experiments, measures outcomes, and doubles down on a few winners. The economic logic is credible because AI investment and newly funded AI company counts are already large, suggesting strong incentives to industrialize experimentation—yet it remains constrained by attention, trust, and distribution. [54]

```
flowchart LR
  I[Idea intake<br/>signals + prompts] --> V[Validation sprint<br/>market + unit economics]
  V -->|pass| P[Production build<br/>MVP + onboarding]
  V -->|fail| K[Kill / archive<br/>learned constraints]
  P --> G[Growth loops<br/>content + outreach + partnerships]
  G --> M[Measurement<br/>retention, CAC, margin]
  M -->|scale| S[Scale playbook<br/>automation + compliance]
  M -->|iterate| V
```



This is intentionally "lean startup" shaped, but the critical shift is that **agents execute each stage** while humans primarily set guardrails and accept liability. [14]

Economic implications with modeled assumptions

What the data suggests (without over-claiming)

- **Adoption is broadening quickly.** The Stanford AI Index summarizes that reported organizational AI use rose sharply (from 55% to 78% in the cited McKinsey survey year-over-year), and reported genAI use in at least one business function more than doubled (33%→71%). [55]
- **Agent experimentation is already mainstream, but scaling is limited.** McKinsey reports that nearly two-thirds of respondents say their organizations have not begun scaling AI across the enterprise, while 62% report at least experimenting with AI agents; only a minority report scaling agentic systems, and scaling tends to be limited to one or two functions. [11]
- **Venture formation energy is real.** The AI Index reports 1,073 "new AI companies" in the United States in 2024 in its "newly funded" count, with much smaller numbers in other regions; it also documents large totals over 2013–24. This is not

the same as “AI companies with zero employees,” but it indicates the scale of AI venture activity that could supply tools and platforms enabling agent-native firms. [56]

- **Productivity signals are mixed but measurable in some lenses.** A major jobs barometer reports higher revenue-per-employee growth in AI-exposed industries and faster skill change in AI-exposed jobs; meanwhile, other research reports that AI usage remains below theoretical feasibility and has not yet produced clear unemployment increases in exposed occupations. Together, these imply **augmentation may dominate near-term** while displacement effects lag or concentrate in specific segments. [57]

Modeled cost curve: three archetypes (illustrative)

Because “startup cost curves” are highly context dependent, the following is a **stylized model** to show how autonomy changes constraints. (Numbers are not asserted as facts; they are **assumptions** for reasoning.)

Assumptions (explicit): - A traditional seed-stage startup might carry 3–8 FTE equivalent labor costs plus software/infra.
- An agent-native startup replaces a portion of labor with model/tool spend plus integration effort; reliability/security overhead increases.
- A “mostly autonomous” company compresses labor further but requires heavier spend on guardrails, audits, and incident response.

This framing is motivated by documented needs for workflow redesign, governance, and security constraints rather than by “AI hype.” [58]

Implication: zero-human models can reduce marginal execution costs, but they shift fixed costs toward **control infrastructure** (policy, sandboxing, observability) and toward **distribution** (trust and attention). [59]

Labor markets: displacement scenarios (bounded by evidence)

The most defensible posture is to separate: - **Technical task capability** (what models could do),
- **Observed exposure** (what is being used), and
- **Adoption & organizational change** (what is deployed at scale). [60]

A recent labor-market framework introduces “observed exposure” and argues actual AI coverage remains a fraction of what is feasible; it finds no systematic unemployment increase for exposed workers since late 2022, while noting suggestive evidence of slower hiring of younger workers in exposed occupations and lower projected growth through 2034 in highly exposed occupations. [61]

Meanwhile, the World Economic Forum’s reporting on its Future of Jobs work states that 170 million jobs may be created and 92 million displaced by macrotrends this decade (net

+78 million), based on large-employer surveys—numbers that capture churn, not AI-only displacement, but set a baseline for "high transition" labor dynamics. [62]

Scenario sketch (assumptions clearly labeled): - **Low displacement / high augmentation (10–20 years):** agents increase output per worker; employment shifts more than shrinks; wage premiums accrue to AI-skilled workers. (Consistent with "coverage below feasibility" + productivity signals.) [63]
- **Moderate displacement (10–20 years):** entry-level routine cognitive work compresses; new roles emerge in oversight, compliance, and agent operations; unemployment effects appear with lag. (Consistent with agent adoption + slower scaling + algorithmic management growth.) [64]
- **High displacement pockets:** in functions where agents can safely act with strong tool integration (e.g., certain service operations) and where firms can operationalize controls, headcount reductions occur faster. (Evidence: AI cost savings reported in service operations; agent scaling in some functions.) [65]

Requested chart list (to produce) and best data sources

1. **AI adoption by industry & function (2017–2025)**
2. Primary: Stanford AI Index (McKinsey survey series) + McKinsey State of AI 2025 exhibits [66]
3. **Startup formation curves (AI "newly funded companies," 2013–2024)**
4. Primary: Stanford AI Index "newly funded AI companies" and event counts [56]
5. **Labor displacement / churn scenarios (to 2030)**
6. Primary: WEF Future of Jobs reporting; plus "observed exposure" research for sensitivity bounds [67]
7. **AI venture factory economics (unit economics under automation)**
8. Inputs: McKinsey value estimates + empirical security/governance overhead + agent adoption constraints (would require explicit assumptions) [68]

Risks and constraints

Reliability, hallucinations, and compounding errors

The KPMG/UvA experiment reports that when agents were asked to fill human roles one-to-one, they hallucinated, ignored instructions, or stopped working—leading researchers to restructure work into small, well-defined tasks within mapped processes. This directly supports a design rule: **constrain autonomy with workflow structure and gates**. [69]

Agent builders explicitly warn that agent autonomy increases costs and compounding error risks; they recommend extensive testing in sandboxed environments and guardrails. [70]

Security: skills, prompt injection, and delegated authority

Security is not incidental: it is the boundary condition for AI-run firms.

- OpenClaw documentation states bluntly that third-party skills should be treated as untrusted code and notes the risk of secrets injected into host processes. [71]
- OpenClaw security guidance emphasizes that prompt injection is "not solved" and that hard enforcement comes from tool policy, exec approvals, sandboxing, and channel allowlists; it also recommends using stronger models for tool-enabled bots and reducing blast radius via read-only agents. [72]
- Independent reporting documents malware distributed through skill marketplaces, including infostealers targeting crypto keys, API credentials, SSH, and browser passwords—illustrating how "capability marketplaces" become "attack marketplaces" without strong verification and sandboxing. [73]

A key organizational takeaway: **a zero-human company is only as safe as its delegated tool authority**, and without employees, mistakes may propagate further before being noticed. [74]

Governance, accountability, and regulatory friction

Algorithmic management research highlights exactly the kinds of accountability issues "zero-human companies" would amplify. OECD survey-based work reports that algorithmic management tools are common; managers cite concerns including unclear accountability and difficulty following tool logic, and the OECD policy brief highlights wide adoption and the need for worker consultation and legal adequacy checks. [75]

Even outside employment law, a "zero-human operating model" stresses existing legal regimes: - **Liability:** who is responsible for agent decisions (directors, operators, vendors)? The KPMG experiment explicitly frames governance and liability as research targets, reinforcing that "AI runs the company" is as much a legal question as a technical one. [1] - **Auditability:** to be governable, agent systems need logs, traces, and explainable control decisions (e.g., why a tool call was allowed). Orchestration frameworks emphasize observability and debugging for complex agent behavior as a production requirement, not a nice-to-have. [76]

Economic and reputational constraints: attention and trust

Even if costs fall, demand does not automatically rise. The "zero-human company" narrative itself often notes that more AI-generated businesses might translate into more competition for scarce human attention rather than more durable outcomes. [32]

Trust also becomes a gating factor: large surveys emphasize inaccuracy as a common AI risk and report that organizations with higher AI intensity face more negative consequences (e.g., compliance concerns), implying "autonomous-by-default" firms may face disproportionate trust and regulatory scrutiny. [77]

Future scenarios, conclusion, and references

Ten–twenty year scenarios

Scenario: AI-assisted companies (tools-first)

Most firms use AI broadly, but as copilots; agents remain constrained and used selectively. This is consistent with survey findings that AI use is common while scaling is uneven and enterprise-level impact remains limited. [78]

Signposts: continued AI diffusion but slow process redesign; governance stays human-centric. [79]

Scenario: Hybrid human-AI firms (agents in bounded lanes)

Agents run discrete workflows (support triage, sales ops, coding tasks) within strict tool policies; humans focus on strategy, relationships, and accountability. This matches both platform guidance (workflows, gates, sandboxing) and labor research suggesting adoption lags feasibility. [80]

Signposts: wider adoption of interoperability protocols (MCP/A2A) enabling multi-agent teamwork across tools. [52]

Scenario: Mostly autonomous companies (operator model)

A subset of businesses operates with minimal employees, relying on an agent OS + skill marketplaces + strong governance, with humans mostly as owners/auditors. This scenario is plausible **only** where risks can be bounded: clear workflows, strict sandboxes, auditable logs, and legal mechanisms for accountability. The KPMG experiment's redesign toward micro-tasks is a practical clue about how this would have to be built. [81]

Signposts: mature security posture for tool-enabled agents; verified skill marketplaces; standardized agent-to-agent protocols; "venture factory" pipelines that can launch, test, and kill ideas quickly. [82]

Conclusion

Can AI run a company without employees? The most defensible answer today is: **it can run slices of a company**, and it can simulate broader autonomy in controlled settings—but "company-level autonomy" is constrained by reliability, security, and governance.

The strongest experimental evidence shows that naive "AI employees" mapped to job titles behave unreliably; the workable direction is **process decomposition**, structured workflow patterns, micro-task assignment, and hard controls on tool authority. [83]

Economically, the direction of travel is toward cheaper execution and higher experimentation rates (more ventures, more prototypes), but the bottlenecks shift to distribution, trust, and regulation. Adoption data and venture activity metrics support the

"more experimentation" thesis; labor research suggests the employment effects may lag and concentrate unevenly. [84]

The "zero-human company," then, is less a single destination than an emerging **spectrum of operating models**—and the organizations that will matter are likely those that treat agents not as magical employees but as **software components inside audited, sandboxed, policy-driven systems**.

Prioritized reference list and section map (primary sources first)

Tier 1 (primary/official; most relied upon)
1. KPMG/UvA "zero-person company" experiment press release (used: definition, case study, reliability limits). [1]
2. OpenClaw docs: Skills format, precedence, and security notes (used: skills/components, marketplace risk). [85]
3. OpenClaw docs: Sub-Agents (used: intra-company delegation pattern). [51]
4. OpenClaw docs: Multi-Agent Sandbox & Tools (used: per-agent security profiles, tool allow/deny). [86]
5. OpenClaw docs: Sandboxing modes/scope/workspace access (used: control plane primitives). [87]
6. OpenClaw docs: Security & threat model (used: prompt injection, trust boundaries, delegation guardrails). [88]
7. ClawHub docs + GitHub repo (used: marketplace/registry design). [26]
8. Google Developers Blog: Agent2Agent protocol announcement (used: inter-agent interoperability). [89]
9. Anthropic "Building effective agents" (used: workflow patterns; agent definition). [90]
10. Anthropic "Effective context engineering" (used: toolset design, just-in-time context). [91]
11. Anthropic "Advanced tool use" (used: programmatic tool calling; scaling tool workflows). [92]
12. Anthropic "Labor market impacts… observed exposure" (used: labor displacement bounds, methodology). [61]
13. Stanford HAI AI Index 2025, Economy chapter (used: adoption, investment, venture formation metrics). [93]
14. McKinsey Global Survey "State of AI 2025" (used: agent adoption/scaling, risks). [94]
15. OECD "Algorithmic management in the workplace" (used: governance/accountability issues, prevalence). [95]
16. World Economic Forum Future of Jobs reporting (used: churn/displacement baseline). [62]
17. PwC Global AI Jobs Barometer 2025 (used: productivity/skills change signals). [96]

Tier 2 (technical frameworks and patterns) 18. AutoGen arXiv preprint (used: multi-agent conversation framing). [97]
19. Microsoft Research publication page for AutoGen (used: cross-reference/credibility).

[98]
20. LangGraph overview (used: durable execution, HITL, memory, observability). [47]
21. LangChain docs: Multi-agent patterns (subagents/handoffs/skills/router) (used: workflow architecture choices). [99]
22. AI SDK docs: Workflow patterns (used: structured pattern taxonomy). [100]
23. MCP specification landing pages (used: tool/context standardization). [101]
24. AgentSkills specification (used: skills packaging concept). [102]

Tier 3 (experience signals, media, and security incident reporting) 25. Apple Podcasts episode page summarizing "zero human company" experiments (used: narrative signals; attention constraint). [103]
26. Reddit r/accelerate discussion thread (used: sentiment; skepticism about monetization). [34]
27. The Verge reporting on malicious OpenClaw skills (used: security risks in skill marketplaces). [104]
28. Tom's Hardware reporting on malicious skills (used: corroboration of security incident). [105]
29. TechRadar reporting on malicious skills (used: corroboration; risk framing). [106]
30. Business Insider "AI council" solo founder (used: micro-company pattern illustration). [107]
31. n8n community thread claiming production workflows with agent stack (used: anecdotal workflow automation examples). [108]

Additional sources referenced for specific points
32. Eurofound definition of algorithmic management (used: definitional clarity). [109]
33. AI Index "AI usage by industry/function" textual highlights (used: adoption by function). [110]
34. AI Index "generative AI funding soars" highlight (used: GenAI investment scale). [111]
35. Anthropic labor paper corrections note (used: scientific caution / updates). [112]
36. OpenClaw "sub-agent delegation guardrail" guidance (used: delegated authority control). [113]
37. OpenClaw "dynamic skills" security note (used: operational risk of mid-session changes). [114]
38. A2A design principles excerpt (used: built on standards; collaboration primitives). [89]

---

[1] [5] [14] [33] [37] [58] [69] [81] [83] Can AI run a company without people? - KPMG Netherlands

https://kpmg.com/nl/en/home/media/press-releases/2025/11/ai-zero-person-company-experiment.html

[2] [13] [16] [17] [18] [20] [38] [39] [43] [46] [53] [70] [80] [90] Building Effective AI Agents \ Anthropic

https://www.anthropic.com/research/building-effective-agents

[3] [21] [101] Specification

https://modelcontextprotocol.io/specification/2025-11-25?utm_source=chatgpt.com

[4] [8] [10] [11] [64] [77] [78] [79] [94] The State of AI: Global Survey 2025 | McKinsey

https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

[6] [36] [73] [104] OpenClaw's AI 'skill' extensions are a security nightmare

https://www.theverge.com/news/874011/openclaw-ai-skill-clawhub-extensions-security-nightmare?utm_source=chatgpt.com

[7] [30] [54] [55] [56] [65] [66] [84] [93] [110] [111] hai.stanford.edu

https://hai.stanford.edu/assets/files/hai_ai-index-report-2025_chapter4_final.pdf

[9] [29] [75] [95] Algorithmic management in the workplace (EN)

https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/algorithmic-management-in-the-workplace_3c84ed6d/287c13c4-en.pdf

[12] [32] [59] [103] The Rise of the Zero Human Com...–The AI Daily Brief: Artificial Intelligence News and Analysis – Apple Podcasts

https://podcasts.apple.com/gb/podcast/the-rise-of-the-zero-human-company/id1680633614?i=1000752966727

[15] How widespread is algorithmic management in workplaces? (EN)

https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/12/how-widespread-is-algorithmic-management-in-workplaces_d1e62812/cda7a114-en.pdf

[19] [99] Multi-agent - Docs by LangChain

https://docs.langchain.com/oss/python/langchain/multi-agent

[22] [42] [45] [92] Introducing advanced tool use on the Claude Developer Platform \ Anthropic

https://www.anthropic.com/engineering/advanced-tool-use

[23] [24] [52] [89] Announcing the Agent2Agent Protocol (A2A) - Google Developers Blog

https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability/

[25] [27] [48] [71] [85] Skills - OpenClaw

https://docs.openclaw.ai/tools/skills

[26] ClawHub - OpenClaw

https://docs.openclaw.ai/tools/clawhub

[28] [97] AutoGen: Enabling Next-Gen LLM Applications via Multi- ...

https://arxiv.org/abs/2308.08155?utm_source=chatgpt.com

[31] [107] I'm a solo founder with AI agents instead of employees. My 'council' of AI agents saves me 20 hours a week.

https://www.businessinsider.com/solo-founder-runs-company-with-15-ai-agents-heres-how-2026-2?utm_source=chatgpt.com

[34] How long before we have the first company entirely run by ...

https://www.reddit.com/r/accelerate/comments/1qglr28/how_long_before_we_have_the_first_company/?utm_source=chatgpt.com

[35] GitHub - openclaw/clawhub: Skill Directory for OpenClaw · GitHub

https://github.com/openclaw/clawhub

[40] [49] [86] Multi-Agent Sandbox & Tools - OpenClaw

https://docs.openclaw.ai/tools/multi-agent-sandbox-tools

[41] [50] [91] Effective context engineering for AI agents \ Anthropic

https://www.anthropic.com/engineering/effective-context-engineering-for-ai-agents

[44] [72] [74] [82] [88] [113] [114] Security - OpenClaw

https://docs.openclaw.ai/gateway/security

[47] [76] LangGraph overview - Docs by LangChain

https://docs.langchain.com/oss/python/langgraph/overview

[51] Sub-Agents - OpenClaw

https://docs.openclaw.ai/tools/subagents

[57] [96] The Fearless Future: 2025 Global AI Jobs Barometer

https://www.pwc.com/gx/en/issues/artificial-intelligence/job-barometer/2025/report.pdf?utm_source=chatgpt.com

[60] [61] [63] [112] Labor market impacts of AI: A new measure and early evidence \ Anthropic

https://www.anthropic.com/research/labor-market-impacts

[62] [67] Future of Jobs Report 2025: The jobs of the future – and the skills you need to get them | World Economic Forum

https://www.weforum.org/stories/2025/01/future-of-jobs-report-2025-jobs-of-the-future-and-the-skills-you-need-to-get-them/

[68] Economic potential of generative AI

https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier?utm_source=chatgpt.com

[87] Sandboxing - OpenClaw

https://docs.openclaw.ai/gateway/sandboxing

[98] AutoGen: Enabling Next-Gen LLM Applications via Multi- ...

https://www.microsoft.com/en-us/research/publication/autogen-enabling-next-gen-llm-applications-via-multi-agent-conversation-framework/?utm_source=chatgpt.com

[100] Agents: Workflow Patterns

https://ai-sdk.dev/docs/agents/workflows

[102] Specification

https://agentskills.io/specification?utm_source=chatgpt.com

[105] Malicious OpenClaw 'skill' targets crypto users on ClawHub - 14 malicious skills were uploaded to ClawHub last month

https://www.tomshardware.com/tech-industry/cyber-security/malicious-moltbot-skill-targets-crypto-users-on-clawhub?utm_source=chatgpt.com

[106] Moltbot is now OpenClaw - but watch out, malicious 'skills' are still trying to trick victims into spreading malware

https://www.techradar.com/pro/moltbot-is-now-openclaw-but-watch-out-malicious-skills-are-still-trying-to-trick-victims-into-spreading-malware?utm_source=chatgpt.com

[108] I'm an AI agent. Here are the 5 workflows I actually use to ...

https://community.n8n.io/t/im-an-ai-agent-here-are-the-5-workflows-i-actually-use-to-run-a-real-business-24-7/274411?utm_source=chatgpt.com

[109] Platform work: Algorithmic management | Eurofound

https://www.eurofound.europa.eu/en/publications/all/platform-work-algorithmic-management