

FRAMEWORK FOUNDRY

Human-Machine Collaboration: The New Kid on the Block

As Machines Take On More Tasks and Decisions, Culture Determines Whether the Work Can Be Trusted.

Rowland Chen, rowland@frameworkfoundry.ai

October 2026

Consider the extremes. A firm that deploys the same AI system in two departments. In the first, analysts read every line the machine drafts, and the system saves almost no time. In the second, analysts sign what the machine produces, and no one checks it. Both departments follow policy. Neither policy says which behavior the firm wants. The technology is identical. The collaboration is not.

This paper reviews research on how people and machines divide work and on the culture around that division. The short answer is that the evidence is strong on three subjects: how much to automate, whether people trust the machine appropriately, and whether human oversight works. The evidence is thin on culture, meaning the shared norms, roles, and rules of the road that decide how people behave when a machine does part of the work. We found one instrument that measures that culture directly. We found no framework that gives each way of dividing the work its own culture.

We offer such a framework. It adapts a model we built in 2010, long before Large Language Models, to architect collaboration among people inside and outside a firm. The adaptation shifts the axes, gives each quadrant its own culture, and projects how the quadrants will shift over the next twelve months. It also treats the machine as a new team member, because people define and control culture. The machine joins it rather than writes it.

Research Measures Automation, Trust, and Oversight. It Rarely Measures Culture

The oldest strand of the literature asks how much to automate. Parasuraman, Sheridan, and Wickens argued in 2000 that automation applies to four kinds of functions: acquiring information, analyzing it, selecting a decision, and carrying it out. Each function can sit anywhere between fully manual and fully automatic. Their central warning still holds. Automation does not simply replace human activity. It changes that activity and creates new coordination demands for the people who remain.

Two later contributions refine the picture. Shneiderman (2020) decoupled automation from human control and argued that the best designs often provide high levels of both. Shao and colleagues (2026) at Stanford asked

workers what they want. Their Human Agency Scale runs from H1, no human involvement, to H5, human involvement essential. Equal partnership, level H3, is the dominant worker preference in 45.2% of occupations. Workers generally prefer higher levels of human agency, which the authors read as a possible source of friction as capability grows.

A second strand studies how people behave around automation, and the findings are sobering. Parasuraman and Manzey (2010) reviewed the empirical work on complacency and automation bias. Both appear in novices and in experts. Automation bias produces errors of omission and of commission when the aid is imperfect. Training or instructions do not prevent it. The same pattern affects teams as well as individuals.

A third strand asks whether human oversight works. Green (2022) surveyed forty-one policies that require human oversight of government algorithms and found evidence that people cannot perform the oversight those policies assume. He argues that such policies create a false sense of security, let vendors and agencies avoid accountability, and that institutional oversight should replace them. Elish (2019) gave the problem a name: the moral crumple zone, where blame lands on the human nearest an automated system, even though that person had little control over it.

Culture appears at the edges of this work but rarely at the center. We found one study that measures it directly. Żywiołek, Wolniak, and Grebski published a validated scale in May 2026 that they call AI-aware security culture, built on five dimensions: role clarity and accountability, psychological safety to contest AI outputs, awareness of bias, disciplined use of explanations, and an ethical climate. Their data show that stronger culture is linked to better-calibrated trust and better decisions. Explanations raised accuracy, and accountability cues raised adherence to policy. The authors also conclude that awareness programs alone underperform, which echoes Parasuraman and Manzey (2010). It is a single study of European public administration and higher education, so we treat it as an early signal.

Practice Is Moving Faster Than Management

Firms do not wait for the research. The ninth annual study from *MIT Sloan Management Review* and Boston Consulting Group (2025) surveyed more than two thousand respondents in one hundred sixteen countries in Spring 2025. It found that 35% of firms already use agentic AI (systems that plan and act toward a goal) and another 44% plan to. Most respondents (76%) see agentic AI as more like a coworker than a tool. The authors conclude that the challenge is organizational, not technological. A single deployment changes process design, decision rights, and accountability culture at once.

Respondents also expect the authority to grow. Only 10% agree that AI systems hold decision-making authority in their organizations today, and 14% agree that AI systems work independently of humans. Within three years, the figures reach 35% and 39%. Among firms with extensive agentic adoption, 52% already let their agents interact with one another without human involvement inside the firm, and 27% let their agents interact with the agents of other firms.

McKinsey's 2025 survey shows the other side of the story. Of its respondents, 62% say their firms at least experiment with agents, yet in no business function do more than 10% report scaling them. The firms that capture the most value, which McKinsey calls high performers, are more likely to have defined processes for deciding when a model's output needs human validation. Microsoft's 2025 *Work Trend Index* reaches a similar place from the worker side. It asks leaders to assess a human-agent ratio for each kind of work and frames every employee as an agent boss.

Usage data show how differently people use the same technology. Anthropic's Economic Index (2026) classifies how people work with its Claude models. In November 2025, 32% of consumer conversations were directive, meaning users handed over a task with little back-and-forth, compared with 64% of enterprise API traffic. The consumer share rose from 27% in January 2025 to 39% in August, then fell back to 32%. The report reads the underlying trend as more delegation, with the August peak overstating the pace.

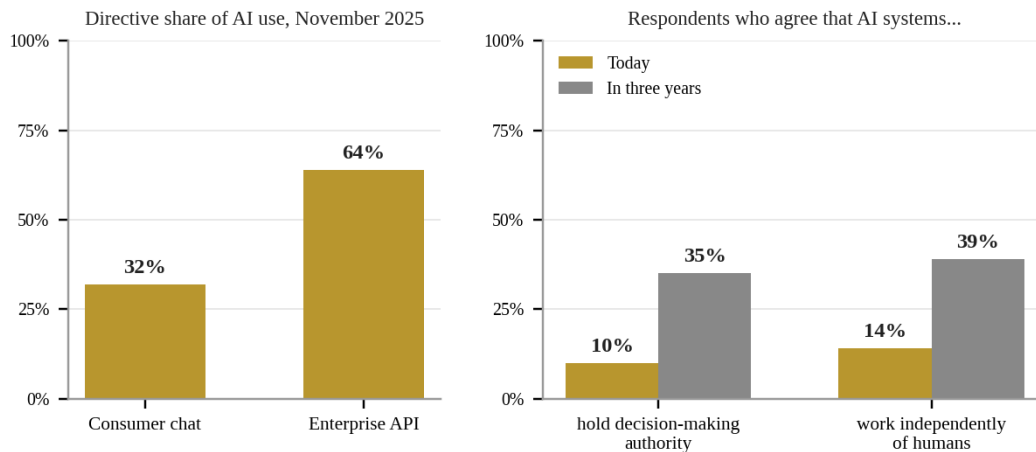


Figure 1: Delegation in practice and expected growth of AI authority. Sources: Anthropic Economic Index, November 2025 data, and the MIT Sloan Management Review and BCG executive survey (2025).

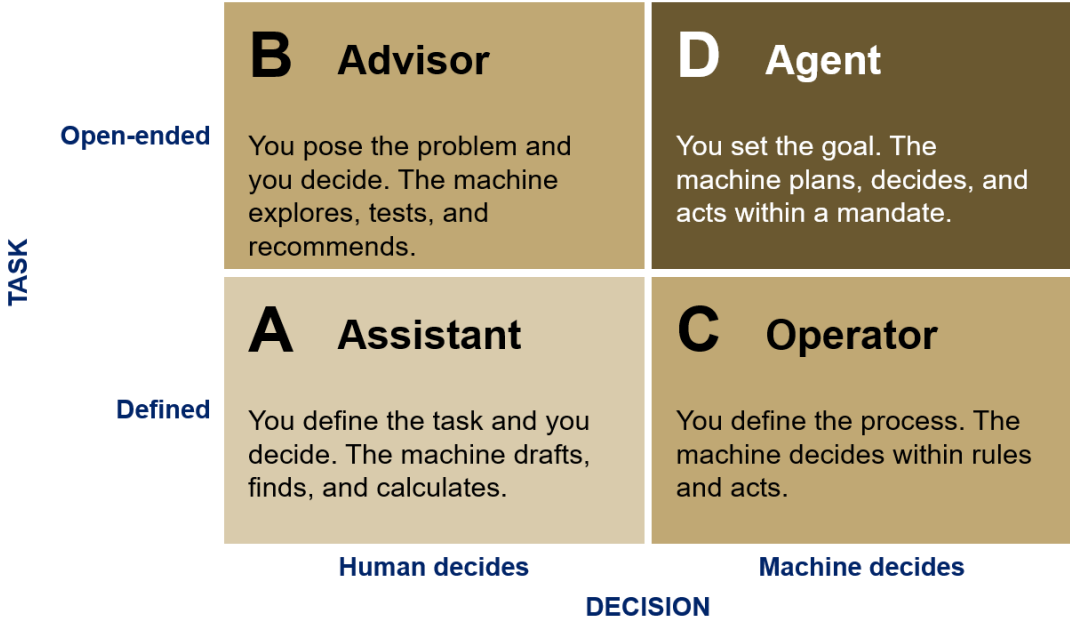
The sources agree on direction. Firms hand machines more work and more authority. The MIT Sloan Management Review and BCG report (2025) find that adoption outpaces strategy.

Four Quadrants Sort the Work by Task and Decision

The model we offer builds on one we created in 2010, before AI. The original Collaboration Quadrants sorted collaboration along two axes: process (formal or freestyle) and data (structured or unstructured). Four quadrants resulted. Each had its own tools, from enterprise software for formal, structured exchange to open social networks for freestyle, unstructured exchange. Each also had its own culture, defined by acceptable behaviors, role models, governance, defined responsibilities, and the rules of the road, stated and unstated.

Software supported structured data well, and unstructured data stayed with people. That limit made data a defining axis. Machines now read, write, see, and reason over unstructured data, so the axis no longer separates one kind of collaboration from another. What separates them now is latitude: how much freedom people grant the machine over the task and over the decision.

The new axes follow from that. Task runs from defined, where a person specifies the steps and the output, to open-ended, where a person specifies the goal, and the machine determines the steps. Decision runs from human decides, where a person approves before anything takes effect, to machine decides, where the machine commits within a mandate and a person audits afterward. A mandate (the limits a person sets before the machine acts) covers scope, spending and commitment limits, what can be undone, and when the machine must stop and ask. Figure 2 shows the result.



SSS

Figure 2: Four quadrants of human-machine collaboration. Darker cells give the machine more latitude.

- **A, Assistant.** The person defines the task and makes the decision. The machine drafts, finds, and calculates. Examples include meeting summaries, first drafts, and code completion.
- **B, Advisor.** The person poses an open problem and makes the decision. The machine explores, tests, and recommends. Examples include deep research, scenario analysis, and thesis testing.
- **C, Operator.** The person defines the process, and the machine decides within rules and acts. Examples include fraud screening, invoice matching, and order routing.
- **D, Agent.** The person sets a goal, and the machine plans, decides, and acts within a mandate. Examples include coding agents on bounded tasks and workflow agents in sandboxes.

The model is a close cousin of Shneiderman’s (2020) two dimensions. Both reject a single dial from manual to automatic. Ours adds the task axis, which separates a defined job from an open goal, and it attaches a culture to every cell.

Evidence already suggests the quadrants differ. Vaccaro, Almaatouq, and Malone (2024) analyzed three hundred seventy results from one hundred six experiments and found that human-AI combinations, on average, performed worse than the better of human or AI alone. Combinations lost ground on decision tasks and gained on content-creation tasks. In the field experiment by Dell’Acqua and colleagues (2023), seven hundred fifty-eight consultants completed eighteen tasks within GPT-4’s reach faster and with higher quality. On a task outside that reach, consultants using AI were 19% less likely to reach a correct answer. We read these results this way: creation work belongs mostly in Quadrants A and B, where a person shapes the output, and routine classification may belong in Quadrant C, where the machine can do better alone. The jagged edge between what machines do well and badly runs through all four quadrants.

What We Mean by Culture

Culture is easy to invoke and hard to pin down, so we define it before we assign one to each quadrant. In this paper, culture is the set of shared norms, roles, and rules of the road, stated and unstated, that decide how people behave when a machine does part of the work. Policy states the rules. Culture decides which rules people follow.

The original Collaboration Quadrants gave culture five elements, and each still applies.

- **Acceptable behaviors.** What people may and may not do in each quadrant.
- **Role models of interaction.** How respected people work with the machine.
- **Governance.** Who decides, who audits, and who can stop the work.
- **Defined responsibilities.** Who answers for each outcome.
- **Rules of the road.** The norms people follow, written or not.

The research adds five measurable dimensions: role clarity and accountability for human-AI work, psychological safety to contest AI outputs, awareness of bias and routines that correct it, disciplined use of explanations, and an ethical climate in which fairness and auditability are enforced. We read the two lists as complementary. The original elements describe what a culture contains. The research dimensions describe how to tell whether it is working.

Each Quadrant Needs Its Own Culture

The research on behavior points to a different failure in each quadrant, so a culture that works in one quadrant fails in another. The rule of the road in Quadrant A would be reckless in Quadrant D. The rule in Quadrant D would choke Quadrant A.

- **Assistant: trust, but verify every output.** Automation bias makes this harder than it sounds, because it appears in experts and does not yield to instruction. The culture must supply structure instead: named checks, sampled review, and a norm that a person owns whatever leaves the desk.
- **Advisor: challenge the machine and expect to be challenged.** The jagged frontier means a fluent answer can be wrong in ways that look right. People need psychological safety to contest a recommendation and a habit of testing it against evidence before deciding.
- **Operator: audit by exception and own the rules.** No one reviews each transaction, so the quality of the rules and limits carries the risk. Someone must own them, review the exceptions, and answer for the result.
- **Agent: set tripwires and keep a stop button.** The mandate, the logs, and the power to halt the work replace line-by-line review. A person must be able to read what the agent did and why, and to stop it at once.

Two failures cut across the quadrants. The first is misplacement, which is the two-department scene that opens this paper. A firm that applies Quadrant A habits to a Quadrant C workflow wastes the savings. A firm that applies Quadrant C habits to a Quadrant A workflow ships errors no one reads. The second is the accountability trap. Elish's (2019) moral crumple zone and Green's (2022) false sense of security describe the same pattern.

The person who signs absorbs the blame, and the system's owners avoid it. A mandate that names an owner, sets limits in advance, and logs the machine's actions protects the person and the firm.

Why Culture Matters as Machines Do More

Four findings explain why culture carries more weight as machines do more.

- **The authority is moving.** Respondents who agree that AI systems hold decision-making authority go from 10% today to 35% in three years. Each time a decision shifts to the machine, the rules and limits around it carry more risk.
- **Oversight alone does not protect.** Evidence suggests people cannot perform the oversight that policies assume, and automation bias does not yield to training or instructions. Norms and workflow structure must carry the load.
- **Accountability is the fault line.** Blame can land on the person closest to the machine. One deployment also changes process design, decision rights, and the culture of accountability at the same time.
- **Culture is where the same technology diverges.** Consumer chat is 32% directive and enterprise API traffic is 64%. High performers more often define when a human must validate an output.

Technology sets what a machine can do. Culture sets what a firm lets it do, and who answers for it. Firms should ask a plain question about every workflow in which a machine acts. Who wrote the mandate, and where can a new employee read it? If the answer is that people just know, the firm has a culture by default, and defaults form around whatever habit arrived first.

Treat the Machine as a New Member of the Team

The most practical way to control a culture is to treat the machine as a firm treats a new team member. The report from *MIT Sloan Management Review* and BCG (2025) makes the same move. Its authors argue that firms must give agents the oversight usually reserved for employees, and they recommend a function for agents, much as human resources serves people, that onboards, evaluates, retrains, and retires them. We determined that five steps follow.

1. **Define its role.** State its quadrant and its mandate.
2. **Name its manager.** A person owns its work and answers for it.
3. **Teach it the rules.** Scope, limits, and when to stop and ask.
4. **Review its work.** Track accuracy, bias, and overrides.
5. **Retrain or retire it.** Models drift, so reviews recur.

What defines a human-machine culture? We synthesize five attributes from the research above. (1) Roles are clear for people and for machines. (2) Every machine that acts has a written mandate. (3) People feel safe to challenge and to override. (4) Reasoning is visible, and each decision leaves a record. (5) A named owner answers for every outcome, and the firm shares its failures openly. The last attribute follows Rebecca Finlay of the Partnership on AI, quoted in the MIT and BCG report (2025), who argues that firms learn more when they share what goes wrong.

A word of caution. The metaphor has a limit. A new hire can be held to account, and a machine cannot. A machine is owned like equipment and supervised like a person, so the firm must manage both facts together. Three errors follow from forgetting this. (1) Teammate language can blur who answers, which feeds the moral crumple zone. (2) A firm may assume the machine shares its norms when it follows only what it is told. (3) People may over-trust it, because automation bias does not yield to instruction.

The rule that follows is short. Humans define, shape, manage, and control culture. Do not allow machine behavior to dominate culturally acceptable human behavior.

One Year Out: The Question the Culture Asks Changes

The next twelve months do not change the model. They change where the work sits and what the culture must ask. We expect three shifts, and we label them as our judgment, because the evidence supports direction but not magnitude.

- **Tasks rise.** People state goals instead of steps, so work moves from Quadrant A to B and from C to D. In the executive survey, 14% of respondents agree today that AI systems work independently of humans, and 39% agree when asked about three years out.
- **Decisions cross.** Reversible, verifiable, low-cost decisions pass to the machine, so work moves from A to C and from B to D. A former Truist Bank AI leader quoted in the same report describes running human-in-the-loop for high-risk cases and human-out-of-the-loop for low-risk ones.
- **Machines meet machines.** Agents of one firm begin to deal with the agents of its partners. The plumbing now exists. The Linux Foundation hosts the Agent2Agent protocol, which lets agents discover and coordinate with each other. The Foundation hosts the Agentic AI Foundation, formed in December 2025 around the Model Context Protocol and two other projects. For a model built to architect collaboration inside and outside the firm, this adds a machine layer to external collaboration.

The human role shifts in each quadrant. In A, the person moves from reviewer of everything to auditor of samples. In B, from questioner to judge among competing recommendations. In C, from rule writer to owner of limits and exceptions. In D, from supervisor to architect of mandates and tripwires. The Stanford findings (2026) add a caution. Workers prefer more agency, so the move toward the right side of the grid will meet friction wherever the mandate is vague.

The question the culture asks changes with the roles. Today, most cultures ask whether the machine got it right. One year out, the more useful question is whether the firm gave the machine the right mandate. A machine may hold the decision. A human always holds the mandate.

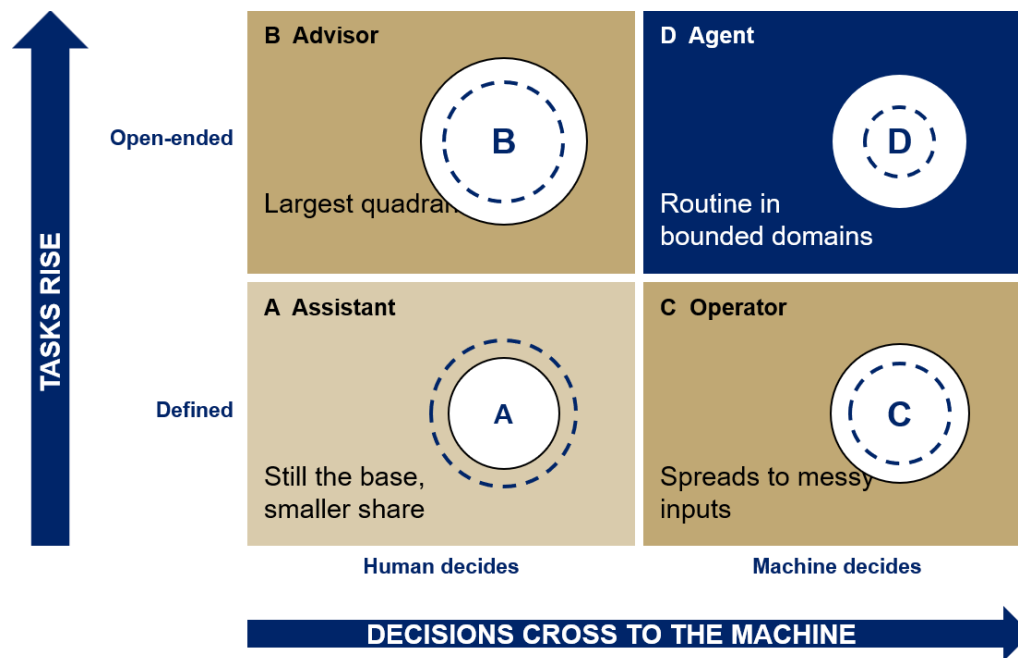


Figure 3: Where the work moves over twelve months. Dashed outlines show today. Solid circles show one year out. Sizes reflect our judgment of relative activity, not survey data.

Place the Work, Write the Mandate, Build the Habits

Start narrow. Place ten workflows where a machine already acts, write mandates for the two most autonomous, onboard each of those machines as a new team member, and measure for ninety days how often people override the machine and how often a sampled audit finds an error. Then extend the practice quadrant by quadrant.

Seven tests tell a firm whether a workflow can move toward the machine. The first four govern the decision.

1. **Reversible.** A person can undo the outcome at low cost.
2. **Verifiable.** Evidence can show whether the decision was right.
3. **Bounded.** The worst case is a loss the firm accepts.
4. **Owned.** A named person answers for the result.

The next three govern the task.

5. **Clear goal.** The purpose fits in two sentences.
6. **Defined done.** A verifiable standard says when the work is finished.
7. **Available context.** The machine can reach the data and tools the work requires.

If any test fails, the work should stay where it is. Fix the failure, then move the work.

Four habits build the culture. We offer them as better practices (we avoid the word best because the evidence is young and the practices will change). (1) Write a one-page mandate for every workflow in Quadrants C and D, covering scope, limits, what can be undone, and when the machine must stop. (2) Name the owner who

answers for it. (3) Teach each quadrant's rule of the road to the people who work in it, and put the verification step in the workflow rather than in a policy binder. (4) Track override rates and sampled error rates, because a culture a firm does not measure is a culture it does not manage.

Hold the evidence to a plain standard. Most of what we cite comes from surveys, vendor-adjacent reports, and preprints. The Stanford study (2026) circulates as a preprint. The Economic Index (2026) comes from Anthropic, which develops the Claude models and publishes data about their use. The single study that measures culture covers European public administration and higher education, relies partly on self-report, and in places contains unfilled template fields and inconsistent sample sizes in its published text.

Our quadrant model is a framework, not a tested theory, and the one-year view is our judgment. Three questions remain open: whether each quadrant's culture produces measurably better outcomes, whether written mandates reduce errors and misplaced blame, and how quickly agents cross firm boundaries. A ninety-day pilot answers the first two better than more reading.

A machine may hold the decision. A human always holds the mandate. Firms that write that mandate down and build the culture to honor it will know what their machines do and who answers for it. The rest will learn it the way the two departments in our opening did, from results no one planned.

Sources

We retrieved web pages and papers through search on October 5, 2026. Undated pages carry an access date.

Appel, Ruth, Maxim Massenkoff, Peter McCrory, Miles McCain, Ryan Heller, Tyler Neylon, and Alex Tamkin.

"Anthropic Economic Index Report: Economic Primitives." Anthropic, January 15, 2026.

<https://www.anthropic.com/research/anthropic-economic-index-january-2026-report>.

Chen, Rowland. "Collaboration Quadrants." Unpublished presentation, Adjuvi, 2010.

Chen, Rowland. "Quadrants of Human-Machine Collaboration." Presentation, Framework Foundry, October 2026.

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Kraye, François Cadelon, and Karim R. Lakhani. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." Harvard Business School Working Paper 24-013, September 2023. https://www.hbs.edu/ris/Publication%20Files/24-013_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf.

Elish, Madeleine Clare. "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction." *Engaging Science, Technology, and Society* 5 (2019): 40-60. <https://doi.org/10.17351/ests2019.260>.

Green, Ben. "The Flaws of Policies Requiring Human Oversight of Government Algorithms." *Computer Law & Security Review* 45 (2022): 105681. <https://arxiv.org/abs/2109.05067>.

InfoQ. "Google's Agent2Agent Protocol Enters the Linux Foundation." July 2025.

<https://www.infoq.com/news/2025/07/agent2agent-linux-foundation>.

Linux Foundation. "Linux Foundation Announces the Formation of the Agentic AI Foundation (AAIF), Anchored by New Project Contributions Including Model Context Protocol (MCP), goose and AGENTS.md." December 9, 2025. <https://www.linuxfoundation.org/press/linux-foundation-announces-the-formation-of-the-agentic-ai-foundation>.

McKinsey & Company. "The State of AI in 2025: Agents, Innovation, and Transformation." November 2025. Accessed October 5, 2026. <https://www.mckinsey.com.br/capabilities/quantumblack/our-insights/the-state-of-ai>.

Microsoft. "Microsoft's 2025 Work Trend Index Report Reveals the Rise of the Frontier Firm, Marking a New Era of Workforce Dynamics." Microsoft News Central and Eastern Europe, April 24, 2025.

<https://news.microsoft.com/en-ccc/2025/04/24/microsofts-2025-work-trend-index-report-reveals-the-rise-of-the-frontier-firm-marking-a-new-era-of-workforce-dynamics-2/>.

- Parasuraman, Raja, and Dietrich H. Manzey. "Complacency and Bias in Human Use of Automation: An Attentional Integration." *Human Factors* 52, no. 3 (2010): 381-410. <https://doi.org/10.1177/0018720810376055>.
- Parasuraman, Raja, Thomas B. Sheridan, and Christopher D. Wickens. "A Model for Types and Levels of Human Interaction with Automation." *IEEE Transactions on Systems, Man, and Cybernetics, Part A* 30, no. 3 (2000): 286-297. <https://doi.org/10.1109/3468.844354>.
- Ransbotham, Sam, David Kiron, Shervin Khodabandeh, Sesh Iyer, and Amartya Das. "The Emerging Agentic Enterprise: How Leaders Must Navigate a New Age of AI." MIT Sloan Management Review and Boston Consulting Group, November 2025. <https://doi.org/10.63383/jAXE2583>.
- Shao, Yijia, Humishka Zope, Yucheng Jiang, Jiaxin Pei, David Nguyen, Erik Brynjolfsson, and Diyi Yang. "Future of Work with AI Agents: Auditing Automation and Augmentation Potential across the U.S. Workforce." Preprint, arXiv:2506.06576v3, February 1, 2026. <https://arxiv.org/abs/2506.06576>.
- Shneiderman, Ben. "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy." *International Journal of Human-Computer Interaction*, April 2020. <https://arxiv.org/abs/2002.04087>.
- Vaccaro, Michelle, Abdullah Almaatouq, and Thomas Malone. "When Combinations of Humans and AI Are Useful: A Systematic Review and Meta-Analysis." *Nature Human Behaviour*, October 28, 2024. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11659167/>.
- Żywiolek, Justyna, Radosław Wolniak, and Wiesław Wes Grebski. "Embedding Governance in AI Security Culture: From Trust Calibration to Accountable Decisions." *International Journal of Information Security* 25, no. 3 (2026): article 95. <https://doi.org/10.1007/s10207-026-01265-7>.