AI SYSTEMS

Augmentation over Automation: People, Process, and Infrastructure as Preconditions for Organizational AI

Why AI succeeds as leverage on human specialists, not as their replacement: the field-experiment evidence, the productivity J-curve, and the capability stack organizations must build bottom-up.

Bhaskar Roy Sarkar, cazyweb · July 17, 2026 ·

AI augmentation · organizational AI adoption · productivity · human-AI collaboration · AI strategy · general-purpose technologies

Download PDF

ABSTRACT

A popular narrative holds that businesses can now be built and run by AI agents, with people reduced to spectators. The empirical record says something more interesting and more demanding. Field experiments consistently find large productivity gains when AI augments skilled workers inside well-understood workflows - and measurable harm when it is trusted outside its jagged capability frontier, including one randomized trial in which experienced developers were slowed down while believing they had sped up. Economic analysis explains the pattern: AI is a general-purpose technology whose returns depend on complementary investments in process, data, infrastructure, and training, producing the familiar J-curve in which measured productivity dips before it climbs. This paper reviews the augmentation evidence, the automation-first

fallacy, and the organizational complements literature, and proposes a practical sequencing model - the capability stack - in which applications precede agents, data structure precedes knowledge tooling, query systems precede retrieval-augmented generation, and explicit process precedes automated skills. We ground the argument in first-party observations from an AI-native agency whose production agents amplify named specialists under human gates. The conclusion is blunt: AI is a part of a business, not the whole of one, and the durable advantage goes to organizations whose people are made harder to compete with by what is built underneath them.

1. Introduction

The loudest claim of the current AI cycle is an automation claim: that agents can replace teams, that software categories are dying weekly, that an entire company can be cloned from a repository and run unattended. The claim is mostly made by people selling something, and it is contradicted by the strongest evidence we have - randomized and quasi-experimental studies of AI in real work - which tells a consistent augmentation story instead: large gains when capable people use AI inside well-understood workflows, real harm when the technology is trusted beyond its uneven frontier, and returns that arrive only after organizations build the unglamorous complements underneath.

This paper reviews that evidence and draws the operational conclusions. Section 2 summarizes the field-experimental record on productivity. Section 3 examines the frontier problem: where AI helps, where it hurts, and why self-assessment fails. Section 4 reviews the economic argument against automation-first framing. Section 5 connects organizational AI to the general-purpose-technology literature and its J-curve. Section 6 proposes a sequencing model for building AI into an organization - the capability stack - and grounds it in our own deployment. Section 7 states the management implications. Our first-party observations are architectural and descriptive; the quantitative claims in this paper all belong to the cited studies.

2. What the Field Experiments Actually Show

The best-identified studies of generative AI at work share a design: real tasks, real workers, randomized or staggered access to the tool. Four results anchor the literature.

In customer support, Brynjolfsson, Li, and Raymond studied a generative AI assistant deployed to thousands of agents and found productivity gains of roughly fifteen percent on average - with the crucial detail that the gains concentrated among newer and less-skilled workers, whose behavior the tool nudged toward that of top performers. The assistant diffused the tacit knowledge of experts; it did not replace the experts whose behavior generated that knowledge.[1]

In professional writing, Noy and Zhang's randomized experiment found ChatGPT cut task time by about forty percent while raising output quality, again compressing the gap between weaker and stronger writers.[2] In software, Peng and colleagues found developers with GitHub Copilot completed a standardized implementation task about fifty-six percent faster than controls.[3]

And in consulting, Dell'Acqua and colleagues ran a field experiment with over seven hundred BCG consultants: for tasks inside the model's competence, AI users completed more tasks, faster, at markedly higher quality. The same study contains the warning that gives it its name, and Section 3 turns to it.[4]

Read together, the pattern is unambiguous. These are augmentation results: the unit of improvement is a person using the tool inside a defined workflow. None of these studies removed the human and measured what the model shipped on its own.

3. The Jagged Frontier, the Ironies of Automation, and the Perception Gap

Dell'Acqua and colleagues describe AI capability as a jagged frontier: tasks of apparently similar difficulty fall on opposite sides of an invisible boundary. Inside it, consultants gained forty percent in quality; on a task deliberately chosen just outside it, consultants using AI were nineteen percentage points less likely to reach the correct answer, because the tool's fluency made its wrong answers persuasive.[4]

This failure mode was predicted four decades earlier. Bainbridge's classic analysis of automation observed that automating the easy majority of a task leaves the human with the hardest residue - supervision, exception handling, recovery - precisely the parts that atrophying skills and misplaced trust make harder to perform.[5] The lesson transfers intact to language-model agents: the more capable the automation appears, the more skilled and engaged the supervising human must be.

The most sobering recent datapoint concerns self-assessment. METR's 2025 randomized trial gave experienced open-source developers early-2025 AI tools on their own repositories and measured a nineteen percent slowdown - while the same developers estimated they had been sped up by twenty percent.[6] Perceived productivity is not productivity. Organizations that roll out AI without measurement are flying on exactly the instrument this study broke.

4. The Automation-First Fallacy

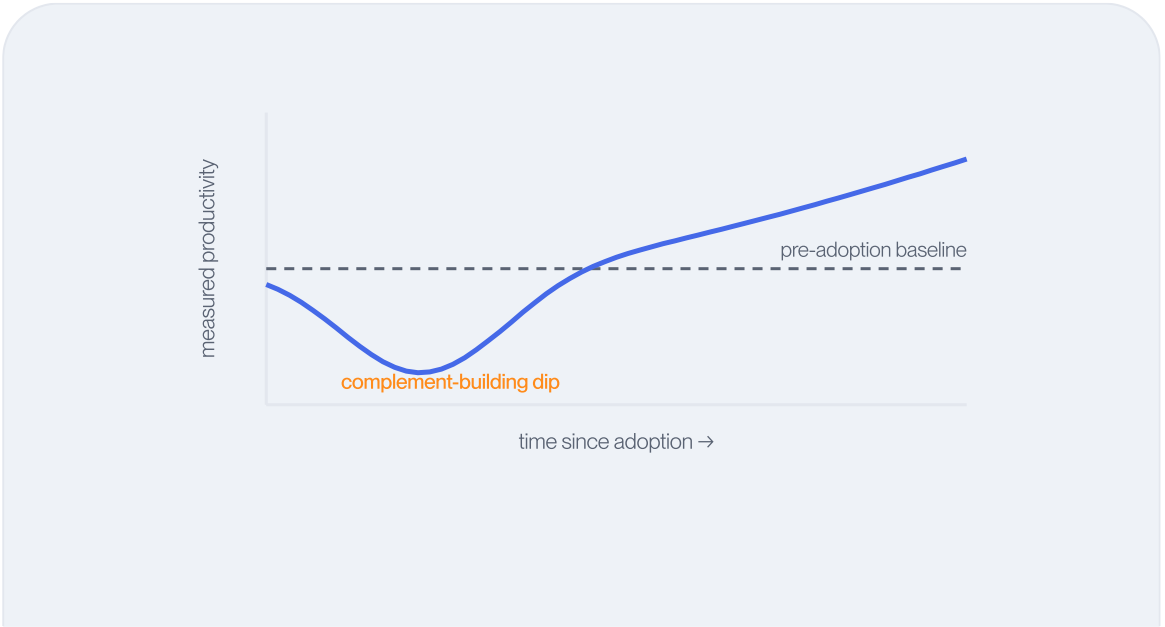
Why does the replace-the-team framing underperform even on its own economic terms? Brynjolfsson's Turing Trap argument: technology built and bought to imitate and substitute for humans competes with labor at existing tasks, where the economic headroom is capped at the cost of the person displaced; technology built to augment humans creates capabilities and tasks that did not exist, where the headroom is uncapped.[7] Acemoglu's task-level analysis reaches a complementary conclusion from the macroeconomic side: given how many tasks AI can actually perform at quality today, automation-driven gains are modest - cumulative total-factor-productivity effects on the order of a percentage point or less over a decade - and the larger prizes

require new tasks and new products, not cheaper substitutes for existing labor.[8]

The practitioner translation is direct. An agent that replaces a mediocre execution of an existing task earns, at best, the wage it displaced, and it inherits the jagged-frontier risk of Section 3 with no human backstop. A specialist whose reach is extended by AI - who ships what used to require a team, at timelines that used to be unrealistic - is playing the uncapped game. The advantage compounds with the specialist, because the specialist is what keeps the output on the right side of the frontier.

5. General-Purpose Technologies and the J-Curve

If AI is so useful, why do so many organizational initiatives show so little? The general-purpose-technology literature answered this for electricity and for IT, and the answer carries over. Brynjolfsson, Rock, and Syverson formalized it as the productivity J-curve: general-purpose technologies demand large, mostly intangible complementary investments - process redesign, data assets, new skills, new organizational forms - which standard measurement books as cost while booking none of the asset. Measured productivity therefore dips first and climbs later, and the depth of the dip scales with how transformative the technology is.[9]



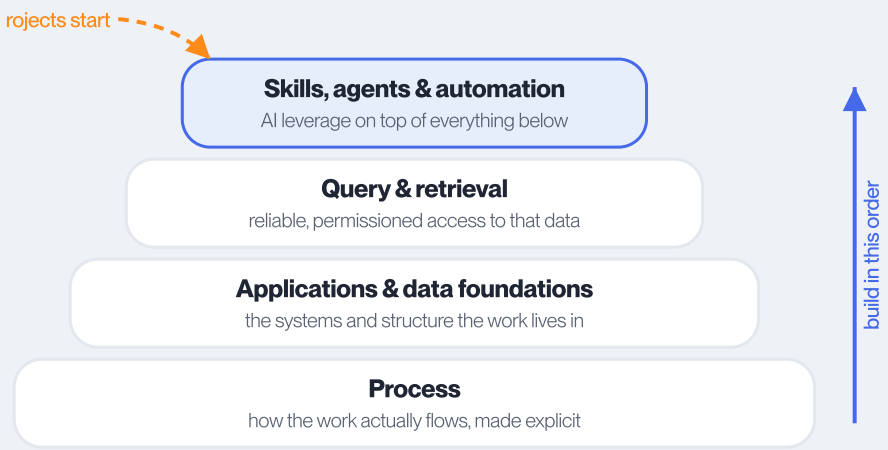
The productivity J-curve: after adopting a general-purpose technology, measured productivity often dips while the organization builds the invisible complements - process redesign, data, training - and rises only once they are in place.

The intangibles are not only technical. Multi-year survey research by MIT Sloan Management Review and BCG finds that the organizations reporting significant benefits from AI are overwhelmingly those that changed how teams work - and that the majority of adopters who treated AI as a drop-in tool saw cultural and process benefits pass them by.[10] Complements are the point, not the overhead.

This is also the honest explanation for the turnkey-agent disappointment cycle. Benchmark-maximizing agents evaluated without cost or reliability constraints overstate real-world readiness[11]; multi-agent frameworks fail predominantly through specification and coordination defects rather than model weakness[12]; and the practitioner guidance from the model builders themselves is to start with the simplest composable pattern and add autonomy only as needed[13]. A downloaded repository contains none of your process, none of your data structure, none of your permissions, and none of your training. It is the top of a stack with nothing underneath it.

6. The Capability Stack: Sequencing Organizational AI

Our operating experience - running an agency on a platform with six production agents embedded in it - compresses into a sequencing rule we have not seen stated plainly in the literature, though it follows from it. Each layer of organizational AI only performs on top of the layer below:



The organizational AI capability stack: agents and automation are the top layer, and each layer only performs on top of the one beneath it. Failed initiatives typically start at the top; durable ones are built bottom-up.

Process before skills. A workflow no one can describe cannot be accelerated. Making the process explicit - stages, owners, quality gates - is the precondition for automating any step of it. This is the intangible investment of Section 5 in its most concrete form.

Applications and data foundations before agents. An agent needs something to act on: systems holding real state, with real permissions and audit trails. Our production agents work because they operate inside an operations platform that already modeled the business - projects, tasks, clients, invoices - with authorization enforced at the data layer. An agent without an application underneath it is a chat window.

Structure before knowledge tooling. Organized file structures, naming conventions, and ownership come before any knowledge vault or memory system delivers value; tooling amplifies structure, it does not create it.

Query systems before retrieval-augmented generation. If a human cannot reliably and safely query the data, a retrieval pipeline will not fix that for a model; it will serve the same ambiguity with more confidence. Our analytics agents answer questions well precisely because they sit on a validated query

layer with a defined metric catalog - the semantics were settled before the model arrived.

Two further observations from operating this stack. First, the leverage lands on specialists: the same agents that make our senior designers and developers dramatically faster produce unusable output when pointed at a task with no specialist owner, which is the jagged frontier of Section 3 experienced from the inside. Second, individual AI fluency did not compose into organizational capability on its own. What one person achieves chatting with a model is a personal workflow; twenty people getting consistent leverage on client work required shared infrastructure, shared frameworks, explicit training, and human gates at the expensive boundaries - engineering and management, not prompting.

7. Implications: People Remain the Operating System

None of this diminishes what AI does; it locates it. A business is a team, a brand, customers, and the relationships among them. The evidence reviewed here says AI multiplies the people in that system - most powerfully the newest ones[1] and the most expert ones inside their domain[4] - and that it cannot substitute for the judgment that keeps work on the right side of the frontier.[5][6] For leaders, the practical consequences:

1. **Buy augmentation, not headcount replacement.** Fund AI where it extends a named specialist inside a defined workflow; treat any fully-autonomous pitch as carrying the burden of proof.[7]
2. **Budget for the dip.** The complements - process mapping, data structure, infrastructure, training - are the investment, and measured productivity may lag while they are built.[9]
3. **Sequence bottom-up.** Process, then applications and data, then query systems, then agents and skills. Initiatives that start at the top of the stack are the ones that populate the failure statistics.

4. **Measure, do not poll.** Perceived speedup and actual speedup can point in opposite directions; instrument the workflow before and after.[6]
5. **Train the frontier.** The scarce organizational skill is knowing where the model is trustworthy; that knowledge lives in people and must be deliberately built.[4][10]

8. Conclusion

The augmentation record is now strong enough to state as an operating principle: AI is a part of a business, not the whole of one. The field experiments show its gains flowing through people; the frontier studies show its failures flowing through unsupervised trust; the economics shows its returns gated on complements that only an organization - not a model, not a repository - can build. The companies that win this cycle will not be the ones that replaced their teams. They will be the ones whose specialists became impossible to compete with, because of the process, the infrastructure, and the training built underneath them - by people, for people.

References

1. Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at Work. *The Quarterly Journal of Economics, 140(2), 889-942*. https://doi.org/10.1093/qje/qjae044
2. Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381(6654), 187-192*. https://doi.org/10.1126/science.adh2586
3. Peng, S., Kalliamvakou, E., Cihon, P., et al. (2023). The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. *arXiv preprint*. https://doi.org/10.48550/arXiv.2302.06590
4. Dell'Acqua, F., McFowland, E., Mollick, E. R., et al. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *Harvard Business School Working Paper No. 24-013*. https://doi.org/10.2139/ssrn.4573321
5. Bainbridge, L. (1983). Ironies of Automation. *Automatica, 19(6), 775-779*. https://doi.org/10.1016/0005-1098(83)90046-8

6. Becker, J., Rush, N., Barnes, E., et al. (2025). Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity. *arXiv preprint (METR)*. https://doi.org/10.48550/arXiv.2507.09089

7. Brynjolfsson, E. (2022). The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence. *Daedalus, 151(2), 272-287*. https://doi.org/10.1162/daed_a_01915

8. Acemoglu, D. (2025). The simple macroeconomics of AI. *Economic Policy, 40(121), 13-58*. https://doi.org/10.1093/epolic/eiae042

9. Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The Productivity J-Curve: How Intangibles Complement General Purpose Technologies. *American Economic Journal: Macroeconomics, 13(1), 333-372*. https://doi.org/10.1257/mac.20180386

10. Ransbotham, S., Candelon, F., Kiron, D., et al. (2021). The Cultural Benefits of Artificial Intelligence in the Enterprise. *MIT Sloan Management Review and Boston Consulting Group*. https://sloanreview.mit.edu/projects/the-cultural-benefits-of-artificial-intelligence-in-the-enterprise/

11. Kapoor, S., Stroebl, B., Siegel, Z. S., et al. (2024). AI Agents That Matter. *arXiv preprint*. https://doi.org/10.48550/arXiv.2407.01502

12. Cemri, M., Pan, M. Z., Yang, S., et al. (2025). Why Do Multi-Agent LLM Systems Fail?. *arXiv preprint*. https://doi.org/10.48550/arXiv.2503.13657

13. Schluntz, E., & Zhang, B. (2024). Building Effective AI Agents. *Anthropic*. https://www.anthropic.com/engineering/building-effective-agents

Cite this

APA Copy

```
Bhaskar Roy Sarkar (2026). Augmentation over Automation: People,
Process, and Infrastructure as Preconditions for Organizational AI.
cazyweb Research. https://cazyweb.com/research/augmentation-over-
automation-organizational-ai
```

BIBTEX Copy

```
@techreport{augmentation-over-automation-organizational-ai,
  author = {Bhaskar Roy Sarkar},
  title = {Augmentation over Automation: People, Process, and
Infrastructure as Preconditions for Organizational AI},
  institution = {cazyweb Research},
  year = {2026},
  url = {https://cazyweb.com/research/augmentation-over-automation-
```

```
organizational-ai}
}
```

RELATED READING

AI will not run your business. People will.

How we run our agency on AI agents

Want a team to run experiments like this for you? See how we help

Let's build something that converts.

Start a project with us, or join the bench of specialists and agencies we deliver with.

Start a project ↗

cazyweb

The engineering and growth team behind leading brands and companies across the US and Europe.

CRO and ecommerce insights

Occasional, genuinely useful notes. No fluff.

you@company.com

Subscribe →

Work email only. No spam. Unsubscribe any time. Privacy.

Solutions

- Product Optimization
- Experimentation Program
- cazyweb Platform

For Ecommerce

- CRO Program
- Email & Retention
- Paid Ads
- Landing Pages
- Revenue Calculator FREE
- A/B Test Calculator FREE
- CRO Agency
- Shopify CRO
- Ecommerce Growth
- Website Performance

Company

- About
- For Brands
- For Agencies
- Careers HIRING
- Contact

Explore

- Services
- Pricing
- Performance program
- Subscription
- Website + CRO from $99
- Case Studies
- Blog
- CRO Ideas
- Apps
- Learn

CazyWeb IT Solutions
GSTIN: 19JTLPS8476P2ZZ
hello@cazyweb.com

© 2026 cazyweb. All rights reserved.