Application

Founders, Agents, and Output - Productivity and Labor Effects of Generative AI

Varsha Shukla, Ph.D. Innovation Fellow, Thayer School of Engineering, Dartmouth College
varsha.shukla.th@dartmouth.edu, +1(603)322-0273, 15 Thayer Dr, Hanover, NH

Graduate Study & Enrollment

Ph.D. Innovation Fellow in Engineering Sciences (Generative AI), Dartmouth College (2022–present; ≥2 years of Ph.D. coursework completed). Advisor: Prof. Geoffrey G. Parker. Expected graduation: June 2026.

Research Interests

Entrepreneurship, labor, and productivity in the digital economy; the startup production function under AI; causal measurement of AI quality and adoption; organizational design (agentic vs. monolithic workflows); governance and audit-grade evaluation standards.

Dissertation

Generative AI and the Startup Production Function: Measuring Founder Productivity, Error Costs, and Labor Demand

Dissertation Summary

Startups are laboratories for technological change: they adopt early, move fast, and pivot when tools disappoint. Yet we know surprisingly little about how generative AI changes the *production function* of a young firm, who does what, how quickly, with what quality, and at what cost. My dissertation begins from a simple premise: measurement precedes management. I construct founder- and team-level measures of AI intensity, the share of core tasks performed with LLMs and agentic tools, by linking prompt/response logs and agent playbooks to the daily "exhaust" of work: code commits and issues, document edit trails, CRM activity, and support tickets.

With these data I ask three questions central to entrepreneurship, labor, and productivity. First, does higher AI intensity raise output per founder-hour once we price error severity and rework? Second, does agentization (a planner coordinating specialized sub-agents) complement managerial and domain capital, reallocating work toward integration roles, or does it substitute away junior generalists? Third, do these organizational shifts translate into faster time-to-MVP, improved sales velocity, and near-term firm survival?

Identification relies on staggered internal rollouts, pre-specified playbooks, and, where feasible, within-team randomization of tasks to agentized vs. baseline processes, estimated with stacked DiD and event-study checks. I also exploit operational frictions (temporary API gates, policy toggles) as instruments for adoption intensity. The objective is evidence that is useful: for

founders deciding how to organize, for workers investing in skills, and for policymakers assessing whether AI diffusion broadens opportunity or quietly narrows it.

Papers / Works in Progress

1. AI Intensity and Founder Output
   Question. Does greater AI intensity raise net throughput once we account for severity-weighted error costs and rework?
   Design. Staggered team rollouts; team × time fixed effects; stacked DiD with pre-trend, placebo, and leave-one-team-out checks.
   Outcomes. Tasks/week, cycle time, PR lead-time, defect severity, time-to-MVP, and sales-funnel conversion.
   Interpretation. Productivity that ignores error is illusory; the test is whether net output rises.
2. Agentized vs. Monolithic Workflows
   Question. Do planner-and-subagent workflows complement managerial and domain capital?
   Design. Randomized task assignment to agentized playbooks; queueing/throughput modeling; heterogeneity by task novelty and operator skill.
   Labor lens. Hiring mix (engineer/analyst/PM), contracting share, wage dispersion. A shift toward "integration" roles indicates complementarity, not blanket substitution.
3. Compliance as Strategy: The Economics of Evaluation
   Question. When do firms choose audit-grade evaluation (accuracy, robustness, fairness, drift), and how does that choice affect adoption and pricing?
   Design. A model of evaluation intensity under regulatory/reputational risk; empirical calibration using time-stamped policy/standards shocks.
   Policy use. If evaluation investments unlock safer deployment in high-stakes contexts, diffusion may broaden rather than concentrate.

Methods, Data, and Ethics

- Data. De-identified operational logs (code/issues, documents, CRM, tickets) linked to AI usage metadata (prompt/response, latency/cost, evaluation flags).
- ID Strategy. Staggered rollouts; pre-registered agent playbooks; within-team randomization where feasible; operational shocks as instruments.
- Robustness. Event-study windows, stacked DiD, saturation with team × product fixed effects; leakage/anticipation checks; severity-weighted error accounting.
- Ethics & Compliance. No PHI; partner IRB/DUA processes where applicable; secure storage; reproducible codebooks.

Policy & Standards Engagement (IEEE-USA)

My research is grounded in practice through active work with IEEE-USA's AI Policy Committee. I contribute to briefings on evaluation, auditability, export controls, and workforce implications, including Congressional Visits Day, where technical concerns are translated into implementable policy levers (measurement requirements, disclosure, procurement standards). In parallel, I collaborate with IEEE working groups on evaluation protocols, accuracy, robustness,

fairness, drift, so that "evaluation intensity" is not an abstraction but a set of operational tests with compliance hooks. This work shapes my empirical design: it clarifies how to price error severity, define risk-adjusted productivity, and exploit policy-timing shocks suitable for causal inference.

Economic Futures, Anthropic Forum

I am invited to participate in Anthropic's Economic Futures Forum (2025), where I advance a framework, "Skill, Not Displacement," arguing that durable gains from generative AI rest on complementarity with managerial and domain capital rather than raw model access. I propose three measurable margins: (i) severity-weighted error costs as a constraint on diffusion; (ii) the rise of integration roles (product/data/operations) as a leading indicator of organizational complementarity; and (iii) evaluation intensity as both a governance tool and a predictor of durable productivity. The Forum provides a venue to pressure-test identification strategies with policy and industry experts, improving external validity and the mapping from firm-level productivity to labor outcomes.

Contribution and Fit with NBER (PIE, IO, Digitization; Labor cross-list)

- Measurement. A field-grounded measure of AI intensity that values quality as well as speed.
- Labor. Evidence on who is complemented (integrators/managers) and who is substituted (junior generalists), with implications for training and wages.
- Entrepreneurship. Whether AI lowers entry costs and broadens participation, or, by raising integration and evaluation requirements, advantages the well-capitalized.
- Policy relevance. Guidance for accelerators, lenders, and workforce programs: where to subsidize integration skills, when to require evaluation, and how to avoid lock-in to brittle tools.

Summary / Abstract

This project measures how generative AI reshapes the startup production function. I assemble founder-level workflow data, code/issue trackers, document edits, CRM and support tickets, linked to AI usage logs to construct task-level AI intensity. Exploiting staggered rollouts of agentic tools and, where feasible, within-team randomization to agentized vs. baseline playbooks, I estimate causal effects on output per founder-hour, severity-weighted error costs, time-to-MVP, sales velocity, and support resolution. I then trace labor responses, hiring mix, contracting share, wage dispersion, and near-term firm outcomes (revenue, fundraising, survival). The analysis separates speed from quality, studies complementarities with managerial and domain capital, and tests whether AI reallocates work toward integrative roles or substitutes away junior generalists. Identification relies on stacked DiD, event-study pre-trends, and operational shocks as instruments for adoption intensity. Findings speak to PIE (measurement & productivity), IO (quality competition and organizational design), Digitization (IT complements), and Labor, informing policies that encourage diffusion while guarding against brittle adoption and unequal gains.