The Deflationary Effects of Generative AI on Digital Goods and Services

Supply Expansion, Consumer Self-Production, Market Consolidation, and Intelligence as Economic Infrastructure

Milan Cheeks
Independent Researcher
November 19, 2026
Theoretical Economics Research Manuscript · Version 1.0

Keywords: Generative AI; technological deflation; digital economics; market entry; consumer self-production; competition; intelligence infrastructure.

JEL: D21, D43, L11, L86, O33, E31, J24

Abstract

Generative artificial intelligence (AI) may transform the economics of digital production by reducing the cost of cognitive tasks, expanding productive capacity, lowering entry barriers, and enabling consumers to produce substitutes for commercially supplied services. This paper proposes a conditional theory of AI-induced relative-price deflation. It distinguishes five channels: declining unit production costs, increased entry, competitive repricing, consumer self-production, and changing vertical market structure. A partial-equilibrium model demonstrates that prices fall when the price-reducing supply effect exceeds AI-induced demand expansion. A complementary model considers the possibility of downstream commoditization alongside upstream concentration in computing and foundation-model infrastructure. The paper develops falsifiable hypotheses and an empirical research program across software, creative services, customer support, and professional services. It argues that sectoral quality-adjusted price declines need not imply economy-wide deflation. This is a theoretical working paper; no original causal price estimates are claimed.

1. Introduction

Generative AI changes more than the productivity of established firms. It can alter who can produce digital outputs, how quickly competing products enter the market, and whether consumers need to purchase services that they can now generate themselves. These mechanisms may shift effective supply outward while changing the demand for monetized services. The central proposition is that, in sufficiently competitive and standardized digital markets, AI-enabled supply capacity can expand faster than monetized demand, exerting downward pressure on quality-adjusted prices. This proposition is conditional, not a prediction that every service becomes free. Verification, differentiation, compute scarcity, demand expansion, and market power may prevent prices from falling. The study asks when productivity gains pass through to prices, when self-production displaces purchases, and whether concentrated infrastructure providers capture an increasing share of value.

2. Related Literature

Solow (1956) establishes technological progress as a driver of long-run productive capacity. Acemoglu and Restrepo (2019) model automation as task substitution alongside the creation of new tasks. Acemoglu (2025) emphasizes that the aggregate effects of AI depend on the share of economically affected tasks and the achievable cost savings. Noy and Zhang (2023) report an experimental reduction in completion time and improvement in quality for selected writing tasks; these are productivity results, not estimates of market price changes. Brynjolfsson, Li, and Raymond (2025) document productivity gains from AI assistance in customer support, with heterogeneous worker effects. Industrial-organization theory additionally implies that the pass-through from costs to prices depends on demand elasticity, entry, and market power. This paper integrates those mechanisms with consumer self-production and the upstream/downstream structure of AI markets.

3. AI-Assisted Unit Costs

Let the unit cost of a standardized service be c(A) = c0(1 - alpha A) + v(A), where A is effective AI adoption between zero and one, alpha is the automatable cost share, and v(A) represents inference, integration, and verification costs. The derivative dc/dA = -alpha c0 + v'(A). Therefore AI adoption lowers unit cost if v'(A) < alpha c0. This condition does not establish a price reduction. Firms may retain savings as profit, invest in quality, or use lower costs to serve new demand. The cost structure also differs by task: a generated draft may be inexpensive, while auditing its correctness, obtaining rights, and assuming liability may remain costly.

4. Equilibrium Price Effects

Assume inverse demand P_D(Q) = a - bQ and inverse supply P_S(Q,A) = c0 - delta A + dQ, with b,d,delta positive. Equating the two yields Q*(A) = (a - c0 + delta A)/(b+d) and P*(A) = (ad + bc0 - b delta A)/(b+d). Thus dP*/dA = -b delta/(b+d) < 0, while dQ*/dA = delta/(b+d) > 0. Under these assumptions, AI shifts supply outward, reducing price and increasing quantity. Now allow AI to increase demand: P_D(Q,A) = a + gamma A - bQ. The new derivative is dP*/dA = (d gamma - b delta)/(b+d). Prices fall only when b delta > d gamma. This inequality is the paper's central conditional result: a sufficiently strong demand expansion can offset or reverse the deflationary effect of lower costs.

5. Consumer Self-Production

Consumers increasingly have the option to generate basic websites, text, graphics, translations, and other outputs rather than buy them. Define the effective purchase cost as C_B = P + T_B, with transaction cost T_B. Define self-production cost as C_S = C_AI + T_S + V_S, where C_AI is model access or inference, T_S is the value of consumer time, and V_S is verification and correction cost. Substitution is attractive when C_S < C_B and output utility is comparable. Let monetized demand be Q_D(P,A) = [1 - s(A)]D(P) + N(A), where s(A) is the displaced share of conventional purchases and N(A) represents new paid demand. Self-production can therefore reduce recorded transactions even while total consumption of useful outputs rises. Verification-heavy and high-liability tasks are less readily displaced.

6. Entry and Competitive Saturation

Let firm profit be Pi_i = (P - c_i)q_i - F_i. Generative AI can reduce both variable cost c_i and fixed entry cost F_i, encouraging entry when expected profit is positive. The resulting increase in suppliers may compress markups for close substitutes. We define AI-induced competitive saturation as a condition in which the availability of commercially substitutable outputs grows faster than buyers' willingness to purchase differentiated outputs. Potential observable indicators include supplier density, price dispersion, acquisition costs, business turnover, and margins. Entry will not necessarily rise forever: intense competition may cause exit and consolidation. A dynamic model must distinguish early entry from subsequent shakeout.

7. Quality and Differentiation

Nominal prices alone may be misleading when product quality changes. A simplified price-per-quality index is $P_Q = P/q$, where q is a measured quality index. A rigorous empirical study should use comparable service bundles or hedonic quality adjustment. Brand, legal accountability, security, maintenance, domain expertise, and reliability can preserve differentiation. When candidate generation becomes cheap, verification may become the scarce complement. AI could thus lower prices for generic outputs while increasing willingness to pay for certified outcomes. The strongest price declines are predicted for outputs that are standardized, easily compared, and inexpensive to verify.

8. Two-Layer AI Market Structure

The downstream market includes applications, content, software, and services. The upstream market supplies compute, foundation models, cloud infrastructure, and distribution. Downstream entry may become easier while upstream development retains substantial scale economies. Define downstream unit cost $c_D = w_L l + p_I i + c_V$, where $w_L l$ is labor cost, $p_I i$ is purchased inference cost, and c_V is verification and delivery cost. As automation reduces l, inference pricing may become a more important determinant of downstream margins. Concentrated upstream providers could retain some productivity gains through pricing power, although open-weight models, local inference, hardware competition, and interoperability may counteract concentration. The Intelligence Utility Hypothesis proposes that general-purpose inference increasingly becomes a metered infrastructure input; it is an analogy, not a claim that AI markets must resemble regulated utilities.

9. Industry Predictions

Routine software development may experience falling implementation costs while security, maintenance, and complex integration remain valuable. Standardized graphic design and marketing copy may become more abundant, with differentiated creative direction retaining pricing power. Customer support may become more productive, but benefits may appear as higher service quality or margins rather than lower consumer prices. Digital entertainment may face content abundance while attention, distribution, and trusted intellectual property remain scarce. Professional services may automate document preparation but retain regulatory, accountability, and verification costs. These are hypotheses requiring industry-specific measurement rather than established market outcomes.

10. Labor and Distribution

Automation may displace labor from specific tasks while creating complementary work and expanding output. The effect on total employment depends on the scale of substitution, demand response, and task creation. Lower prices can improve real purchasing power, but lower labor income could offset those benefits for affected households. Value may shift among consumers, workers, downstream firms, and upstream infrastructure owners. The distributional effects therefore matter as much as average productivity. A complete welfare analysis would require household heterogeneity, wages, income effects, and the valuation of nonmarket self-produced outputs.

11. Sectoral Versus Aggregate Deflation

A decline in the price of AI-exposed services is not equivalent to economy-wide deflation. For an approximate aggregate index ln P = sum_i w_i ln P_i, aggregate inflation is approximately pi = sum_i w_i pi_i. Digital service prices influence aggregate inflation according to their expenditure weights and offsetting movements elsewhere. Energy, compute, scarce physical inputs, premium services, new consumption categories, and macroeconomic policy can offset sectoral price declines. The theory more strongly predicts relative-price pressure in selected markets than a sustained fall in the general price level.

12. Dynamic Transition

A plausible sequence begins with early adopters retaining productivity gains as margins. Competitive diffusion then reduces the exclusivity of these advantages. Lower barriers allow new suppliers to enter and some consumers to self-produce. Firms subsequently differentiate, bundle, subscribe, consolidate, or exit. Finally, a larger share of value may migrate to infrastructure, verification, and distribution. This is a conceptual sequence rather than a universal forecast. Markets can skip, reverse, or experience several phases at once.

13. Falsifiable Hypotheses

H1: Greater effective AI adoption reduces quality-adjusted unit costs in exposed service categories. H2: Cost reductions are passed through more strongly where competition is greater. H3: AI adoption increases entry in low-capital, low-regulation digital markets. H4: Consumer-facing AI access displaces purchases of some standardized services. H5: Price declines are smaller where verification, trust, and liability costs are high. H6: Concentrated upstream inference markets capture more surplus from AI-dependent downstream firms. H7: Price effects appear earlier in digitized, verifiable tasks than in physical or heavily regulated activities. H8: Early aggregate inflation effects are smaller than relative-price effects within highly exposed sectors.

14. Empirical Research Design

Construct a longitudinal panel of comparable service offerings across website development, design, translation, copywriting, software maintenance, customer support, and digital media. Record transaction prices, output characteristics, independent quality scores, delivery time, supplier entry, adoption intensity, labor hours, and inference costs. Estimate ln(P_it) = alpha_i + lambda_t + beta(Exposure_i x Post_t) + gamma X_it + epsilon_it. A negative beta would be consistent with the hypothesis only if identifying assumptions hold. Examine pretrends, spillovers, staggered adoption, compositional changes, and alternative exposure measures. Prefer verified transaction prices to advertised listings where available. Separate identical-task repetition, quality improvement, and genuine price declines. Use event studies and appropriate modern difference-in-differences estimators for heterogeneous adoption timing. For self-production, combine purchase records with surveys or randomized AI-access interventions. Measure upstream and downstream concentration separately using HHI = sum_j s_j squared.

15. Illustrative Simulation

For inverse demand $P_D = 100 - Q$ and inverse supply $P_S = 20 - 20A + Q$, equilibrium at $A=0$ is $P=60$ and $Q=40$. At $A=0.5$, $P=55$ and $Q=45$. At $A=1$, $P=50$ and $Q=50$. The assumed model thus produces a 16.7 percent price reduction and 25 percent quantity increase between zero and full adoption. These values are entirely illustrative, not observed data or forecasts. Changing the demand response, market power, or cost assumptions can materially alter the results.

16. Alternative Scenarios

Competitive deflation occurs when costs fall and market entry passes savings to consumers. Demand-led expansion occurs when new uses increase willingness to pay faster than capacity grows. Concentrated value capture occurs when dominant suppliers retain margins. Quality-led premiumization yields cheaper generic outputs alongside costly verified services. Compute-constrained inflation occurs when scarce infrastructure raises inference costs. Broad disinflation occurs if sufficiently weighted sectors experience lower inflation, even if the aggregate price level continues rising. These alternatives must be assessed rather than excluded by assumption.

17. Business Model Implications

As routine production becomes abundant, buyers may increasingly pay for validated outcomes rather than production time. Firms may differentiate through trust, integration, security, customer relationships, distribution, and ongoing support. Subscription access can replace one-off transactions, but subscriptions are not immune to competition. Consumer attention and reliable distribution may become relatively scarcer than basic digital content. AI inference may become a recurring operating input, making supplier diversification and cost management strategically important.

18. Policy Implications

Competition analysis should distinguish downstream markets from upstream infrastructure. Workforce transition programs may support complementary skills, though their effectiveness requires evaluation. Consumer welfare assessment should consider both cheaper services and risks from unreliable output, privacy loss, or reduced accountability. Statistical agencies may need improved quality adjustment and methods to estimate the welfare value of self-produced digital services. Interoperability and access to compute could influence whether productivity gains are broadly distributed.

19. Limitations

This is a theoretical research manuscript, not a completed causal study. AI adoption is endogenous, output quality changes rapidly, and advertised prices may differ from transactions. Self-production can lower spending without lowering welfare. New demand may exceed new supply. Infrastructure concentration is not inevitable. Aggregate price effects require general-equilibrium analysis, including labor income, monetary responses, and investment. The illustrative equilibrium model simplifies heterogeneous firms and consumers. No original dataset or statistical significance claim is presented.

20. Future Research

Next steps include a pre-registered panel study of quality-adjusted prices, structural models of firm entry and exit, randomized studies of consumer self-production, and measurement of inference-price pass-through. Additional work should assess international trade, physical goods enabled by robotics, and the potential for skill verification to become a primary cost center. A general-equilibrium extension should integrate labor demand, household income, investment, and monetary policy.

21. Conclusion

Generative AI can simultaneously lower production costs, increase supply capacity, facilitate entry, and permit consumers to produce substitutes for purchased digital services. A simple competitive model predicts declining prices when supply expansion dominates AI-induced demand growth. The outcome is not universal: quality, trust, compute scarcity, demand creation, and market power can offset it. Downstream commoditization may coexist with upstream concentration, redistributing gains among consumers, workers, firms, and infrastructure owners. The most defensible prediction is conditional relative-price deflation in exposed digital markets, not inevitable economy-wide deflation. Rigorous quality-adjusted price data and causal evaluation are needed to establish the scale and timing of these mechanisms.

References

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

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. Journal of Economic Perspectives, 33(2), 3-30. https://doi.org/10.1257/jep.33.2.3

Brynjolfsson, E., Li, D., & Raymond, L. R. (2025). Generative AI at work. Quarterly Journal of Economics, 140(2), 889-942. https://doi.org/10.1093/qje/qjae044

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

Solow, R. M. (1956). A contribution to the theory of economic growth. Quarterly Journal of Economics, 70(1), 65-94. https://doi.org/10.2307/1884513

Appendix A. Research Dataset

Suggested fields: AI adoption share; comparable transaction price; quality-adjusted price; unit production cost; labor hours; inference cost; verification cost; supplier entry and exit; output volume; self-production share; concentration index; customer acquisition cost; customer satisfaction; service reliability; and regulatory exposure. Each variable requires a written operational definition and consistent observation window.

Appendix B. Integrity Statement

All numerical equilibrium examples in this manuscript are simulated from stated assumptions. The cited productivity experiments are not evidence of market-level deflation. No original causal estimates are reported. Before submission, verify citation metadata, choose a target journal or working-paper series, preregister the empirical protocol where appropriate, and report null as well as positive findings.