Brookings Papers
ON ECONOMIC ACTIVITY

BPEA Conference Draft, September 24-25, 2026

The Vanishing Advantage of Specialization: AI, Knowledge Utilization, and the Boundary of the Firm

Luis Garicano (London School of Economics)

The Vanishing Advantage of Specialization: AI, Knowledge Utilization, and the Boundary of the Firm

Luis Garicano
London School of Economics

Draft for the Brookings Papers on Economic Activity – September 2026

Abstract

Specialists enjoy knowledge-based economies of scale because they can spread the fixed cost of acquiring knowledge across many clients. Artificial intelligence reduces the fixed cost and hence particularly benefits those who use this knowledge less often. Expert work is reorganized: the frequent problems move to the firm that owns them and the production of general, verifiable knowledge can move up to the model makers; the specialist is left with the rare problems for which reliable expertise remains costly to acquire. This expansion of the scope of client firms is the opposite prediction from a fall in transaction costs, so the two changes can be told apart. Preliminary evidence from ChatGPT use, corporate legal departments, and freelance markets is consistent with this shift, as are data showing an accelerating fall in the shares of lawyers, accountants, software developers, and management analysts employed in their main specialist industries.

The division of labor economizes on the cost of acquiring knowledge. The reason is that the specialist acquires knowledge once and utilizes it repeatedly, allowing others to rely on her expertise rather than acquiring it by themselves. Artificial intelligence reduces this fixed cost of acquiring knowledge and therefore also reduces the advantage of specialization.

I study here how this change affects the boundaries of the firm. But the underlying mechanism is more general: cheaper knowledge reduces the advantage of specialists vis a vis both individual generalists and organizations. Hence work that used to require multiple specialists can be brought together under the control of one person or firm. At the firm boundary, this mechanism suggests expert work shifts from specialist suppliers towards their clients.

Recent employment patterns are consistent with this shift. Employment in professional, scientific, and technical services firms grew at 0.5 percent per year from November 2022 to July 2026, compared with 2.6 percent per year during the 2010s and 3.4 percent a year from November 2019 to November 2022. It is not because the number of lawyers, accountants, software developers, and management analysts stopped growing– in fact it grew 2.5 percent a year between 2022 and 2025. But of the 337,000 jobs added by the four occupations, 99% (335,000) were added outside their main specialist industries.

A fall in transaction costs would predict a move of the boundary of the firm in the opposite direction. A growing literature explores how artificial intelligence changes organizational boundaries through its impact on transaction costs. If agentic AI can find suppliers and allow the transaction to take place cheaply and automatically between the agents, the argument goes, within-the-firm transactions could be replaced by market purchases from specialist providers. This movement towards the market would be reinforced by tokenization and programmable contracting. In the limit, the firm could disappear.[1] [2] AI then creates a race between Smith and Coase: cheaper knowledge weakens the Smithian gains from specialization; cheaper transactions the Coasian rationale for the firm. Which force dominates determines whether work moves inside or outside the firm.

A firm is organized around a core activity, such as manufacturing a product, delivering a service, or serving a particular client base. It has local knowledge about that activity, controls it, and is responsible for the results (Knight, 1921). To perform that core activity successfully, many support functions must be supplied, such as law, accounting, software, and marketing.

Knowledge is costly to acquire but, once acquired, cheap to reutilize (Rosen, 1983; Romer, 1990), so that its average cost falls with its utilization. Firms and markets are different mechanisms to increase utilization: inside the firm, experts direct those who do not know (Garicano, 2000); in the market, the specialist sells the same knowledge to many clients, allowing it to be used more intensively (Garicano and Hubbard, 2007). A firm therefore handles frequent problems internally and buys expertise outside for problems it encounters too rarely to justify maintaining the knowledge in-house.

For instance, Microsoft has a large legal department, yet it hired the corporate law firm Simpson Thacher to advise it in its purchase of Activision Blizzard (Chen, 2022). It is not because Microsoft lacks lawyers that it relies on outside service providers, but because it does not use sufficiently often the specialized knowledge of mergers to justify an entire department.

There are many occupations at this make-or-buy margin. Just the six occupations in Table 1 employ 5.9 million workers. In five of the six, most workers are employed outside the specialist industry associated with the occupation. The specialist industries generated $1.79 trillion in receipts, according to the 2022 Economic Census.

Note that the fall in the cost of acquiring knowledge did not begin with generative AI. Google search and databases had already lowered these costs. Generative AI greatly accelerates this process.

Suppose a capability costs $100,000 a year to maintain, a firm faces the problem 10 times a year while the specialist sees it 1000 times, and briefing an outsider costs $2,000 per case. Since the specialist's advantage due to utilization is $9,900 per case, the firm buys the service from the

[1]This is what Shahidi et al. (2025) call the Coasian singularity, after Coase (1937).

[2]Among the reasons for concentrating authority over the core activity are bearing risk and making the best decisions (Knight, 1921); monitoring team production (Alchian and Demsetz, 1972); and using asset ownership to align incentives for actions that are not contractible (Grossman and Hart, 1986).

Table 1. Selected specialist functions: employment location, its change since 2022, and market size

	Employment	In main specialist industry (%)		Growth 2022–25 (%/year)		Receipts
Function	2025 (thousands)	2022	2025	Main specialist industry	Other employers	($ billions)
Legal	754.5	61.4	59.4	1.0	4.0	374.6
Accounting	1,449.5	24.2	22.8	−0.9	1.7	208.8
Software	1,687.9	34.3	30.5	−0.7	5.1	600.6
Marketing	899.6	4.9	4.6	1.9	4.2	134.8
Management analysis	898.3	26.6	24.9	1.2	4.4	435.0
Design	197.8	11.1	11.9	0.0	−2.6	36.8
Six functions	5,887.6	–	–	–	–	1,790.6

Notes: Occupations are Lawyers; Accountants and auditors; Software developers; Market research analysts and marketing specialists; Management analysts; and Graphic designers. The corresponding main specialist industries are Legal services (NAICS 5411); Accounting, tax preparation, bookkeeping, payroll (NAICS 5412); Computer systems design and related (NAICS 5415); Advertising, public relations, and related (NAICS 5418); Management, scientific, and technical consulting (NAICS 5416); and Specialized design services (NAICS 5414). Other employers are user industries, government, and adjacent specialist suppliers. Growth rates are for the number of people in the occupation who work in the main specialist industry and for the number who work for any other employer, in percent a year between May 2022 and May 2025. The 2022 and 2025 figures use the same BLS method; Appendix Tables A1 and A2 give the annual shares since 2015, and the growth rates in earlier periods. Receipts cover all establishments in the listed industry, not only the row occupation. Sources: author's calculations from BLS Occupational Employment and Wage Statistics, May 2022 and May 2025, and the U.S. Census Bureau's 2022 Economic Census, table EC2200BASIC.

specialist. Now suppose AI reduces the annual cost of maintaining the capability to $10,000 for both. The utilization advantage of the specialist is reduced to $990, less than the cost of asking for help across the boundary, so the firm prefers to acquire the capability itself.

We are in the very early stages of AI's impact, but we present some preliminary evidence that appears consistent with this mechanism. First, recent evidence on the breadth of AI use: among work-related ChatGPT messages associated with a particular occupation, 43.5 percent concern tasks of another occupation (Chin and Richmond, 2026). Second, expectations about the firm boundary: corporate legal departments report large increases in generative-AI use and expect to rely less on outside advisers (Association of Corporate Counsel and Everlaw, 2025). Third, the freelance market: after ChatGPT's release, freelancers in writing-related occupations experienced fewer jobs and lower monthly earnings, and the remaining postings asked for more skills (Hui et al., 2024; Demirci et al., 2025). Fourth, the shift in the employment of lawyers, accountants, software developers, and management analysts toward employers outside their main specialist industries since 2022 (Table 1), accelerating a change that was already visible before the pandemic (and consistent with an accelerating reduction in knowledge acquisition costs).

The idea that knowledge utilization drives specialization and organization has been studied before. Garicano and Hubbard (2007, 2009) show that lawyers specialize more in thicker markets, and that utilization, more than firm size itself, determines what remains inside the firm. Ide and Talamàs (2025) introduce AI into knowledge hierarchies and show how autonomous AI handles routine problems while humans handle exceptions, with more knowledgeable humans gaining most. Within a particular job, Garicano et al. (2026) ask when AI unbundles the tasks inside a job. While those papers take the boundary between the client and the specialist as given and study the division of labour among specialists or inside the firm, here the boundary is the object of study.

I. Theory

The firm owns the local knowledge of its core activity. This includes the shared knowledge of facts, rules and language inside the activity, what Crémer (1993) calls its "culture". It also faces a stream of problems that require specialized knowledge in order to be solved. Order the problem classes the firm faces, indexed by z, from common to rare, so that a higher z denotes a problem the firm faces less often. For each problem class, the firm chooses whether to acquire the capability or send the problem to an outside specialist. In what follows, "routine" means common and "exceptional" means rare.[3]

Throughout, we study problems for which a person or organization must take responsibility. That is, even when AI performs most of the research and analysis, someone must implement it or sign for it. The firm must choose whether it handles the problem under its own control or refers it to an outside specialist. We call the first outcome in-house and the second outsourced, regardless of how much of the underlying work is performed by AI. The owner of the core activity chooses where each problem class should be solved to minimize the cost of solving it.[4]

Let $q(z) > 0$ be the number of times the firm faces problem class z over the period the knowledge is useful. By the ordering just defined, $q'(z) < 0$. Let an outside specialist pool n cases of each class across clients, with $n > q(z)$. We take n to be common across classes, so as to keep frequency as the only source of the ordering. Define:

$$u(z) \equiv \frac{1}{q(z)} - \frac{1}{n}. \tag{1}$$

[3]Following Garicano (2000), the same index can instead be read in terms of difficulty. Hold problem frequencies fixed and let the cost of learning increase with z. The baseline analysis below uses the frequency interpretation.

[4]We assume the value of a solved problem is the same regardless of where it is solved, so the cost minimization is without loss, and that acquiring knowledge for one class neither lowers the cost of another nor creates a fixed cost shared across classes, so that we can decompose it class by class. If the capabilities are shared, then the process we describe would take place in discrete bundles.

Because $q(z)$ is decreasing in z, $u'(z) > 0$. $u(z)$ measures the specialist's economy of scale due to knowledge utilization: how much more widely she spreads the same fixed cost. Notice that the relevant primitive is the frequency of a problem class within a firm, not total firm size: a corporation with millions of transactions may experience one large merger a decade.

Let $k(m) > 0$ be the fixed cost, per period and per problem class, of having reliable knowledge when AI is at level m: the AI system plus the fixed cost of establishing and maintaining the ability to review and approve its output. The time spent reviewing each case is a variable cost, assumed equal on both sides. AI may cut the human part sharply; what matters for the model is who holds the knowledge and controls its use. We assume $k'(m) < 0$, common across classes, and extend the analysis below to the case where AI lowers k by different amounts in different classes.

Let $\tau > 0$ be the per-case cost of asking a specialist for help outside the firm's boundary. This requires, for instance, translating the firm's language (Crémer et al., 2007) or explaining the local context.

Hence buying from a competitive specialist costs:

$$\frac{k(m)}{n} + \tau \tag{2}$$

per use, while producing inside costs $k(m)/q(z)$ per use. Note that a price that is charged per API call or per specific usage of the AI is not a fixed cost, and cancels out here, since both parties pay equally for it. What is relevant for the analysis is the part of capability cost that remains fixed, and is divided by q or n.[5]

The firm keeps problem class z inside when

$$\frac{k(m)}{q(z)} \leq \frac{k(m)}{n} + \tau, \qquad \text{or equivalently} \qquad u(z)k(m) \leq \tau. \tag{3}$$

A reduction of k by 1 dollar saves the firm that insources $1/q(z)$ and the firm that outsources the capability $1/n$ per use. Hence the specialist's cost advantage decreases by $u(z)$, *even though both sides use the same technology.* The specialist still spreads the cost over more uses; the saving from doing so is less likely to cover the cost of crossing the boundary.

Note that the same logic that applies to the division of labor between firms also applies to specialization within firms. We only need to interpret $q(z)$ as the number of times that a particular employee encounters a problem of class z and n as the number of times such a problem is handled by

[5] Note that we are studying a competitive benchmark where the specialist gets zero profit at the operating scale n, so the specialist charges average capability cost $k(m)/n$ and the buyer also bears τ.

a colleague or department. τ is then the cost of referring the problem to someone else. A decrease in k expands the range of tasks that an employee can undertake on their own.[6]

Minimizing cost class by class gives a cutoff because $u(z)$ increases as problems get more unusual. At an interior cutoff $z^*(m)$, the costs of making and buying are equal[7]:

$$u\big(z^*(m)\big)k(m) = \tau. \tag{4}$$

The firm handles classes $z \leq z^*(m)$ and buys classes $z > z^*(m)$ from specialists.[8]

Proposition 1 (The vanishing advantage of specialization). A fall in k, holding all else equal, increases the $z^*(m)$ cutoff so that the firm deals inside with a wider range of problems. The outside specialist deals with a rarer set of problems.

The result follows directly from (4). Differentiating with respect to k gives

$$\frac{dz^*}{dk} = -\frac{u(z^*)}{u'(z^*)k} < 0. \tag{5}$$

Thus a fall in k raises z^*: the firm acquires capabilities that it uses less often.[9]

[6]When matching problems with those who know how to solve them is expensive, these referrals are organized in a knowledge-based hierarchy, as Garicano (2000) shows. A fall in k then means more problems are solved by those who face them first, rather than being passed up the hierarchy.

[7]The cutoff is interior if u is continuous and $u(z)k(m)$ is below τ for some classes and above it for others.

[8]Or, in the difficulty interpretation, the firm then handles the easier classes and buys the harder ones from specialists. A proportional fall in learning costs moves the cutoff toward harder problems.

[9]Under the difficulty interpretation, the same result says that the firm takes on harder problems and that the specialist is left with the hardest tail.

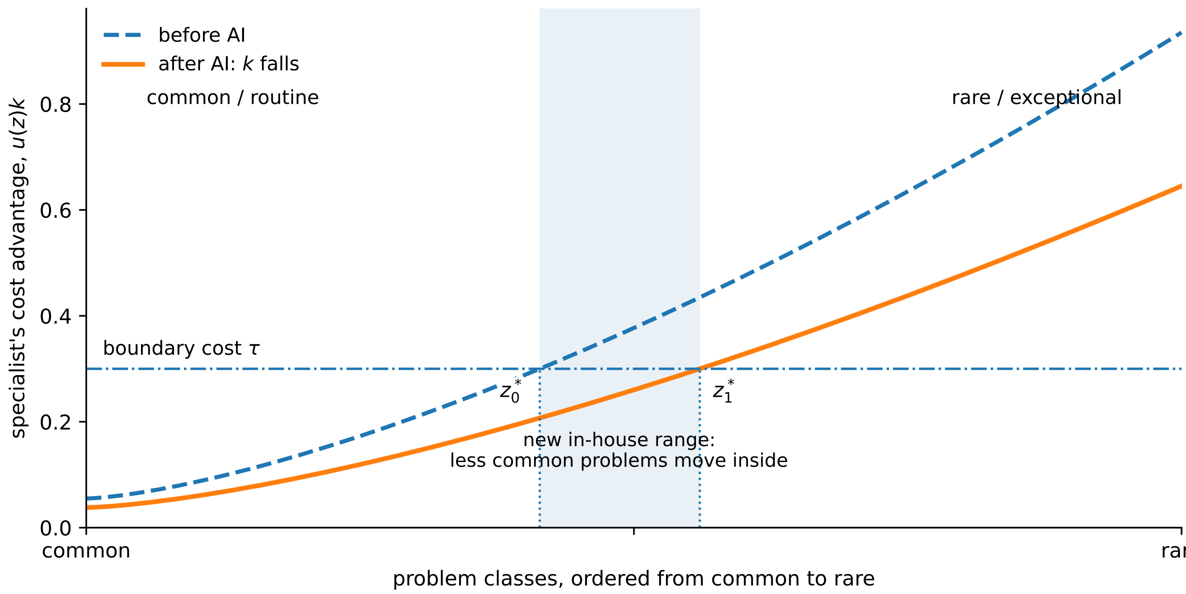

Figure 1. Lower knowledge costs and the firm boundary. Problem classes are ordered from common to rare, so the specialist's advantage from knowledge utilization, $u(z)k(m)$, rises with z. A fall in k reduces the advantage so that the cutoff moves from z_0^* to z_1^*.

Proposition 1 holds the cost of crossing the boundary fixed. AI may also reduce that cost, for example by lowering search and matching costs (Shahidi et al., 2025; Hadfield and Koh, 2025). Let $\tau = \tau(m)$, while holding q and n fixed. A fall in τ moves work outside the firm, the opposite of a fall in k. When both costs fall, z^* rises and work moves inside if, at $u(z^*)k = \tau$,

$$-\frac{k_m}{k} > -\frac{\tau_m}{\tau}. \tag{6}$$

Thus the relevant question is which cost falls faster in proportional terms.

In the other direction, we have the theorem of Stigler (1951) (not formalized by him).

Corollary 1 (Stigler's theorem). Hold k and τ fixed. When market expansion allows an outside specialist to pool more cases, n rises. The specialist's advantage from knowledge utilization increases, and the cutoff moves toward more common problems. Work moves out of firms and into specialist suppliers. The division of labor is limited by the extent of the market.

The evolution of the personal-computer industry is a well-known example of this mechanism: production moved from vertically integrated firms such as IBM to specialists in operating systems, chips, hard drives, graphics cards, and other hardware (Bresnahan and Greenstein, 1999). The same is true for the car industry and its many component suppliers.

The model also gives a prediction across firms. Fix one problem class and let firms differ in the number of times they use it, $q_i(z)$. Firm i makes rather than buys from the outside when

$$q_i(z) \geq \bar{q}(m) \equiv \frac{k(m)n}{k(m) + \tau n}. \tag{7}$$

Here, $\bar{q}$ is the threshold internal use needed to justify owning the capability.

Corollary 2 (Minimum internal use). A fall in k lowers $\bar{q}$ and pushes work inside the firm, while a fall in τ raises $\bar{q}$ and pushes it outside. A rise in n also raises $\bar{q}$ and pushes work out of the firm.

Hence firm size matters only insofar as it raises the frequency with which a particular capability is used. Figure 1 illustrates the results. Once specialist scale is large relative to the marginal client's use, further pooling has little effect: $\bar{q} \approx k/\tau$ when $n \gg \bar{q}$.

General cost comparison.

AI could help specialists by more than their clients. This is not enough to reverse the results, since these savings are spread over more problems. What matters is the saving per problem. To reverse the direction of the result AI must lead to a reduction in the specialist's cost per solved problem by more than the client's.[10]

For an illustration of specialist scale, let one specialist handle at most N cases of a given class per period, and let $Q^M(z)$ be the number of cases of class z sent to the market. Assume average-cost pricing, ignoring indivisibilities, and full utilization in sufficiently large markets, so that a specialist pools

$$n(z) = \min\{N, Q^M(z)\}. \tag{8}$$

This describes specialist scale conditional on the number of cases sent to the market. Under these assumptions, one specialist serves everyone in a small market; in a large one, specialists enter and each fills up. Maintain $n(z) > q(z)$. Market growth raises $n(z)$ while the market is thin; once specialists reach capacity, it raises their number instead. AI can raise N, lowering outside average cost when capacity binds, which is the n_m term in equation (A2).

Problem-specific knowledge costs.

[10] Let k_F and k_S denote the firm's and specialist's fixed capability costs, ℓ_F and ℓ_S denote expected losses from errors per case, and let $\mu \geq 0$ be the markup. The firm then makes when

$$\frac{k_F(z,m)}{q(z)} + \ell_F(z,m) \leq \frac{k_S(z,m)}{n} + \tau(z,m) + \mu(z,m) + \ell_S(z,m).$$

The zero-profit benchmark sets $\mu = 0$. What matters is the saving per case: a larger fall in k_S need not outweigh a smaller fall in k_F if n is sufficiently larger than q. Lower ℓ_S also favors buying. A changing markup moves the boundary like a change in τ.

We have assumed till now that AI lowers the cost of knowledge equally for every problem. In practice, this is not the case. Language models need large amounts of data (Hoffmann et al., 2022; Muennighoff et al., 2023), and their knowledge is easier to deploy when answers can be verified. A problem can be rare for a client, without being rare across clients: a firm may undertake few mergers, but the examples may be plentiful among many firms. The cost of acquiring knowledge can decrease fast when examples are many and answers easily verifiable, like in the case of routine legal cases or code, while it may remain costly where solutions are hard to describe or verify, such as the expertise needed to undertake a one-off reorganization of a company.

Hold q, n, and τ fixed. Allow the fixed cost of reliable knowledge to depend on the problem class, $k = k(z, m)$, and define

$$A(z, m) \equiv u(z)k(z, m) \tag{9}$$

as the specialist's cost advantage from knowledge utilization.

Corollary 3 (Problem-specific knowledge costs). Suppose first that AI weakly lowers the knowledge cost for every problem class, $k_m(z, m) \leq 0$ for all z. Then any class made inside before the improvement remains inside afterward. Some classes may move from buying to making, but no class moves from making to buying. Suppose, in addition, that the specialist's cost advantage rises as problems become rarer:

$$A_z(z, m) = u'(z)k(z, m) + u(z)k_z(z, m) > 0. \tag{10}$$

The firm's choices then continue to have a single common-to-rare cutoff, defined by $A(z^*(m), m) = \tau$, and at an interior cutoff

$$\frac{dz^*}{dm} = -\frac{u(z^*)k_m(z^*, m)}{u'(z^*)k(z^*, m) + u(z^*)k_z(z^*, m)} \geq 0. \tag{11}$$

Thus cheaper knowledge expands the range of problem classes made inside and leaves specialists with the rarer tail. The minimum use that justifies making, $\bar{q}(z, m) = k(z, m)n/[k(z, m) + \tau n]$, weakly falls with m in every class, so Corollary 2 holds class by class as well.

Condition (10) is the single-crossing condition for the general case. It nests the two interpretations discussed above. In the frequency interpretation, the knowledge cost is constant across classes and the ordering comes from $u'(z) > 0$. In the difficulty interpretation, utilization is held fixed and the ordering comes from $k_z(z, m) > 0$.

The corollary establishes two different results. A symmetric fall in knowledge costs always improves the relative advantage of internal provision for that class. Every class of problems handled inside remains inside. The second result is that, under some circumstances, the problems handled by outside specialists become rarer (or harder). The condition is that their cost advantage increases as problems become either rarer or harder (depending on the interpretation). When both frequency and knowledge cost vary, what matters is that $u(z)k(z,m)$, the specialist's total cost advantage, continues to rise from common to rare problems.

We have taken as given the falling capability costs for the specialist and the client firm. A reason for such a fall is that the knowledge is embedded in a model and reused without specialists. The same utilization logic can explain why: the model producer is spreading the fixed cost of developing the knowledge over a very large pool of uses, much larger than that of a professional-service firm. Where access to the model substitutes for acquiring expertise locally, it lowers the fixed cost for both the specialist and the client.

Professional-service firms can be squeezed in the middle between model producers and their clients. The same utilization economies of scale that Corollary 1 relies on can concentrate the production of reusable, general knowledge upstream at the labs, while the application moves towards the owner of the problem. For instance, in the law example, a contract can be designed inside the client firm using the knowledge embodied in the model and the local knowledge embodied in the firm. The professional service firm can then be squeezed between the model producer, where reusable general knowledge is supplied, and the firm below, where it is applied. The specialist maintains an advantage when its expertise remains costly to acquire and clients use it too rarely to justify maintaining it internally.

II. Empirical evidence and measurement

Cheaper knowledge can change who does the work within firms and across their boundaries. An employee may use AI to perform a task previously referred to a specialized colleague. A firm may similarly undertake work previously purchased from outside: its lawyers may draft a contract that would previously have been sent to a law firm. In both cases, cheaper access to knowledge reduces the need to rely on someone who specializes in its use. The evidence below examines these changes by looking at the tasks employees undertake, where specialists are employed, and firms' reliance on outside providers. Before doing that, we first set out the empirical predictions of the mechanism.

II.1. Empirical implications: Smith versus Coase.

Holding utilization and other costs fixed, the Smith and Coase mechanisms lead to opposite predictions here for the direction of the assignment, the mix of outside work, and the internal use that is needed to justify making rather than buying. The table below collects these results in order to guide empirical work.

Table 2. Smith versus Coase: what each theory predicts

Observable	**Trading gets cheaper: τ falls**	**Knowledge gets cheaper: k falls**
Assignment	The firm deals inside with fewer problem classes.	The firm deals inside with more problem classes.
Range of outside work	Specialist suppliers deal with a larger set of problems.	Specialist suppliers deal with a smaller set of problems.
Outside-work mix[a]	Expands to more common problems.	Contracts towards rarer problems.
Minimum internal use	$\bar{q}$ rises. Firms switch from making to buying.	$\bar{q}$ falls. Firms switch from buying to making.

[a] Under the alternative difficulty interpretation described in the text, read "common" and "rare" as "easy" and "hard."

Proposition 1 concerns the assignment of a given flow of problems. This is a reasonable approximation for mergers, whose number is unlikely to respond much to the cost of legal advice. It is less appropriate for compliance checks or lawsuits, whose number may rise when they become cheaper. Let Q_z be the number of problems of class z solved in the economy, s_z the share handled by outside specialists, and $X_z = s_z Q_z$ specialist volume. Then

$$\Delta \log X_z = \Delta \log s_z + \Delta \log Q_z. \tag{12}$$

Specialists can therefore lose share and nevertheless gain volume if the number of problems solved expands sufficiently. The organizational response is also ambiguous. If an individual firm's use of the capability rises, a higher q makes internal provision more attractive. If additional demand thickens the outside market or increases specialist capacity, a higher n strengthens outside provision. Equation (A2) gives the general condition when both respond.

Consider radiology as an illustration of these possibilities. If the cost of reading scans goes down because it is cheaper to interpret them, then demand will increase. Smaller hospitals may choose to develop their own capability to read those scans because their q becomes sufficiently large and there is clinical knowledge about the patient that keeps the transaction cost positive. Outsiders may also benefit from higher demand, but if that is met by new entrants, cases per specialist will not change and so there need not be any further reduction in cost per case from greater utilization. Hence the organizational outcome depends on how utilization inside the hospital changes relative to utilization by the specialist.

Note that both mechanisms can lower the cost of entry. Let q_e denote an entrant's anticipated uses of one ordinary problem class over the period in which the capability is useful. Its minimum total cost for that class is

$$c(q_e, m) = \min\left\{k(m), q_e\left[\frac{k(m)}{n} + \tau\right]\right\}. \tag{13}$$

A fall in k lowers both branches, while a fall in τ lowers the buy branch. Thus entry counts alone do not identify the mechanism, and the entrant's organization still depends on anticipated utilization: it makes when $q_e \geq \bar{q}(m)$ and buys otherwise. Neither mechanism says anything about the upper tail of the firm-size distribution (Appendix A.3).

II.2. Where the specialists work

OEWS reports wage-and-salary employment by occupation and employer industry. We use the share of each occupation employed in its main specialist industry as a suggestive proxy for the specialist share of work; the survey observes neither purchases nor problem assignment and excludes the self-employed. Figure 2 follows the six functions from 2015 to 2025.[11]

From 2015 to 2019, all six occupations grew faster at other employers. The pattern was mixed during the pandemic (2019–2022). The specialist industry shares for lawyers, accountants, software developers, and management analysts fell by 1.4 to 3.8 points after 2022. Lawyers, accountants, and software workers are below their earlier trends; management analysts have returned to theirs. Of the 337,000 jobs created between 2022 and 2025 by the four professions, 335,000 were outside their main specialist industries.

This drop is consistent with the reorganization predicted by falling knowledge costs, but we cannot claim to identify causality here, since this timing fits both the general drop in knowledge costs with the IT revolution, intensified by AI, and the end of the professional-services boom.

[11] Appendix A.4 describes the series, the change in BLS methods in 2021, the survey's three-year panel design, and sampling uncertainty.

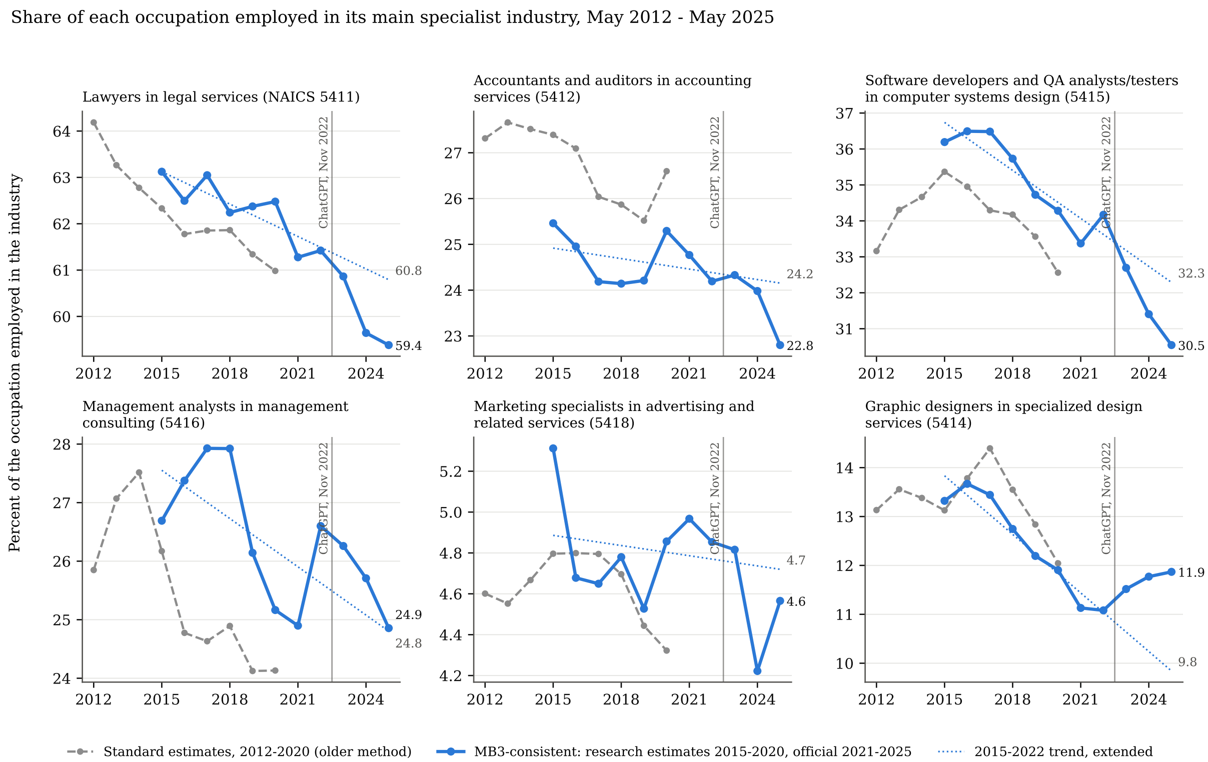

Figure 2. Share of each occupation that works in its main specialist industry, May 2012 to May 2025. Blue: BLS research estimates for 2015–2020 and official estimates for 2021–2025, computed with the same method; the dotted line extends the straight-line trend of 2015–2022. Grey: the older estimates for 2012–2020, computed with the previous method. The vertical line marks the release of ChatGPT in November 2022. Software includes quality-assurance analysts and testers. Sources: author's calculations from the BLS Occupational Employment and Wage Statistics.

II.3. Early evidence on changing boundaries

Specialization and breadth

Recent evidence shows workers using AI to cross their occupational boundaries. Chin and Richmond (2026) classify more than 800,000 work-related ChatGPT messages from users whose occupations are known. Each message is assigned to the occupation whose historical tasks look most similar. Most of the usage is generic, including what one would expect: help in writing, editing and planning represents 61.5 percent of all messages. But 16.8 percent of all work messages, and 43.5 percent of the ones which are specific to one occupation, concern tasks usually undertaken by a different occupation (Figure 3). Among the ones that cross occupational boundaries are calculating financial data, troubleshooting software, drafting promotional messages, and writing to government agencies.

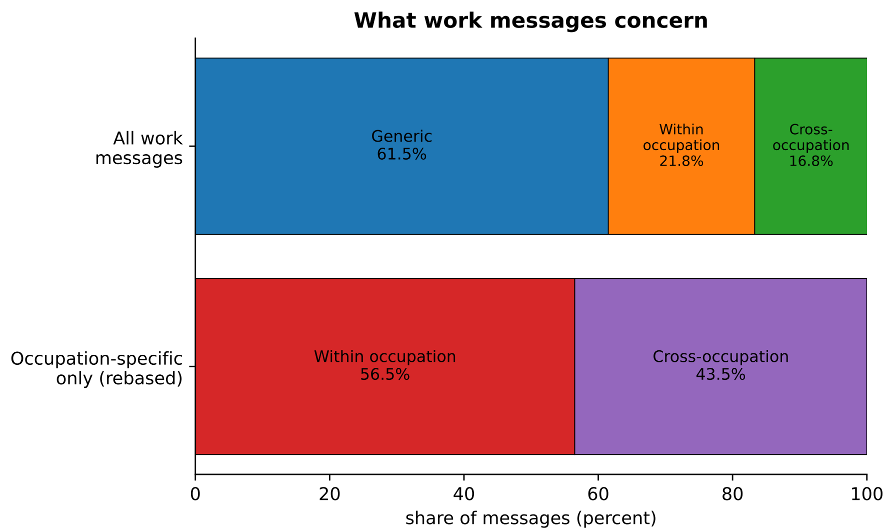

Figure 3. Task crossover in AI use at work. The first bar divides work messages into generic, within-occupation, and cross-occupation tasks. The second drops generic work and rebases the rest. The source reports no standard errors. Source: author's transcription of Chin and Richmond (2026).

Experimental evidence shows that AI can reduce the reliance on other specialists to solve one's own problems. In a field experiment at Procter & Gamble, Dell'Acqua et al. (2026) find that individuals working alone can use AI to produce proposals that are of similar quality to the ones that teams produce without AI. AI allows workers to use knowledge outside their own function, so they can rely less on specialists from other functions.

Ethnographic evidence goes in the same direction. Retkowsky et al. (2026) show that creative workers who are in charge of developing a concept use AI to produce a polished visualization of their own concepts, without relying on other professionals. That visualization, because it comes at a very early stage, constrains the work assigned to others. From the perspective of our model, the originator of the project performs a wider range of tasks in the process.

The example of the law

Early survey evidence from corporate law departments appears consistent with the knowledge channel. The 2025 ACC–Everlaw survey covers 657 in-house legal and legal-operations professionals in 30 countries. Because the 2024 survey covered only the United States, the year-over-year comparisons are restricted to U.S. respondents. Among them, reported use of generative AI in legal practice rose from 23 to 52 percent. Sixty-four percent expect generative AI to reduce reliance on outside providers for routine legal tasks, and 50 percent expect it to reduce the cost of outside providers. In the global 2025 sample, the main tasks in which respondents expect to bring more hours in-house are drafting, contract management and compliance tracking, and legal research

(Association of Corporate Counsel and Everlaw, 2025). These are just expectations from a survey, not purchases.

In the first eleven months of 2025, in a sample of 184 U.S. law firms, billable hours grew 1.9 percent and worked hourly rates 7.3 percent (Thomson Reuters Institute and Georgetown Law, 2026). A smaller share and a larger volume can go together, as we have argued above.

Change in the composition of work of specialists

Freelance platforms already reduce the cost of finding, contracting with, and checking outside help. A further fall in those costs would favor outsourcing. But freelance writers saw 2.0 percent fewer monthly jobs and 5.2 percent lower monthly earnings after ChatGPT's release, relative to freelancers in less-exposed occupations (Hui et al., 2024). Postings in automation-prone writing and coding fell by about one fifth relative to manual-intensive categories (Demirci et al., 2025). Among employers posting both before and after ChatGPT, the remaining automation-prone posts had 5.7 percent higher maximum advertised budgets and 2.2 percent more required skill tags (Demirci et al., 2025).

Note that this is only indicative evidence–none of this is direct proof of the make versus buy margin, since we do not observe the internal production of the buyer, nor the frequency in each class of problems pre-AI. A posting that disappears may mean the buyer now does the job fully with AI, which is an inward move in our sense, or that the job is no longer worth doing at all. Other selection mechanisms can be operating. In the conclusion we discuss the evidence that is needed to differentiate these explanations.

III. AI and the regulation of professional-service firms

There are three margins of policy that determine whether the reallocation that the model predicts, and the potential productivity gains, are in fact allowed to happen: licensing, competition, and training.

Incumbent professional-service firms whose expertise loses value may oppose the falling price of knowledge for their clients. Changes do not only happen at the boundary of the firm, but also between occupations, as those outside a given occupation are able to solve problems for which they previously needed expert help. This reduces potentially the role of experts. Happily for them, professional-service firms are formed by people in professions that can write their own rules. They help decide what falls under the unauthorized practice of law, who can sign an audit, and so on. [12].

A key decision for society is how far we want regulation to make insourcing and across-occupation solving of problems difficult. Where the solution is equally reliable, and there is not a specific reason

[12]The classic study on how professions divide and contest control over expert work is Abbott (1988)

to mandate independence (think of external auditors), rules that constrain the ability of internal professionals to sign their own solutions, including, for instance, restricting the tasks of in-house lawyers or nurse practitioners, can reduce gains and dynamism.

Second, competition decides whether the fall in k reaches the firms that would use it. If the owners of the models, or of the data that train them, have monopoly power, they can charge high fixed access fees or restrict access to their models. This could keep internal provision expensive and weaken the effect that we are discussing. Note the contrast with high per-token charges which do not affect the make versus buy decision in the symmetric benchmark, as they affect both equally.

Finally, training. Professional service firms run an apprenticeship that produces deep expertise for the whole industry. Routine work supports this arrangement in two ways. First, this is how the training gets financed. Juniors pay for their training with cheap labor on routine cases, and the firm recovers the cost by billing clients for that work. Second, routine work supplies the problems that are actually needed for juniors to acquire experience. AI can affect both: it can remove the routine work that financed the arrangement, and with it the material on which juniors learned by doing.

The specialization of firms in harder and less frequent tasks may further limit learning opportunities, as juniors may not have enough experience to contribute to these cases. A countervailing force may be an increase in the value of the experts on these harder problems, which makes the career more valuable and increases the incentive to acquire expertise (Garicano and Rayo, 2025). Lawyers and accountants can still be trained, inside client firms or in schools, but new institutions may have to emerge to provide a training ladder that routine work used to provide.

IV. Conclusion and future work

A specialist earns her living by learning something once and using it for many clients. When anyone with a model can get the same knowledge for a fraction of the price, that advantage shrinks relative to the occasional user, who spreads the cost of learning over a smaller number of problems. So firms do a larger share of their support work themselves and rely on outsiders for a smaller share. The experts can be squeezed from above, by model makers or labs who supply the general knowledge at scale, and from below, by their clients who can apply this cheaper knowledge themselves. The specialist's comparative advantage lies in expertise that clients use too rarely to maintain internally and that remains costly to acquire even with AI, including the expertise required to check and verify a solution. Cheaper knowledge, as we have seen, brings problem solving inside the firm while lower transaction costs lead to buying more solutions outside.

We present some early evidence that points at the mechanism we suggest, but it is early days in this technological revolution. A test of our hypothesis would require focusing on just one kind of work, such as employment lawsuits, recording how often the firm faces it and who takes responsibility for each case before and after the firm gets AI: its own lawyers, a law firm, or nobody. A causal test would also need a comparison that separates AI adoption from changes in demand and the firm's circumstances.

Online Appendix

A.1 The race with all four forces

Let q, n, k, and τ all depend on (z, m). The firm makes class z when

$$G(z,m) \equiv k(z,m)\left[\frac{1}{q(z,m)} - \frac{1}{n(z,m)}\right] - \tau(z,m) \leq 0, \tag{A1}$$

and a marginal improvement favors making when

$$u(z,m)\left[-k_m(z,m)\right] + \frac{k(z,m)q_m(z,m)}{q(z,m)^2} > -\tau_m(z,m) + \frac{k(z,m)n_m(z,m)}{n(z,m)^2}, \tag{A2}$$

with $u(z,m) = 1/q(z,m) - 1/n(z,m)$. On the left hand side, cheaper knowledge and higher utilization within the firm increase what the firm makes, the problems it deals with. On the right hand side, lower transaction costs andmore cases per specialist increase what the firm buys. If $G_z > 0$ there is still a single cutoff, and at an interior cutoff (A2) gives $dz^*/dm > 0$. With q and n fixed, (A2) reduces to $u(z)[-k_m] > -\tau_m$. At the cutoff $u(z^*)k = \tau$, so dividing by τ gives the proportional condition (6) in the text.

A.2 Workspace-size comparison

For users in the middle half of message volume, Chin and Richmond (2026) report cross-occupation shares of 18.9, 17.9, 17.9, 17.7, and 16.3 percent for workspaces with 2–5, 6–15, 16–25, 26–100, and 101 or more seats: a 2.6-point gradient with no standard errors, which disappears among heavy users. Seats measure the ChatGPT workspace, not the company, so we do not use the gradient as evidence on firm size.

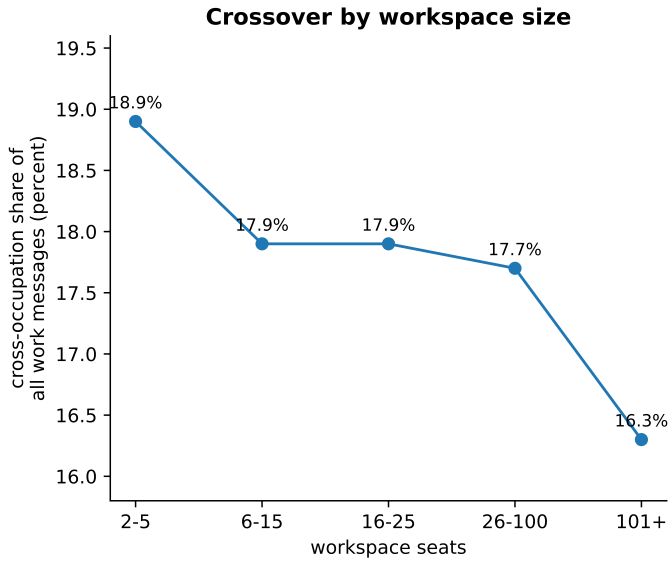

Figure A1. Cross-occupation share of work messages by workspace seats, users in the middle half of message volume. No standard errors are reported. Source: author's transcription of Chin and Richmond (2026).

A.3 Entry and the employer-firm size distribution

Entry is a margin on which the knowledge- and transaction-cost explanations can coincide: either can reduce entry costs, though the make–buy choice still depends on anticipated utilization. Employer startups increased by 12.3 percent from 2019 to 2023 while employment per startup dropped 11.1 percent (U.S. Census Bureau, 2025). The monthly average of seasonally adjusted Business Applications in 2025 was 60.3 percent above its 2019 average. Applications with planned wages rose only 2.9 percent between 2019 and 2025 (U.S. Census Bureau, 2026). Decker and Haltiwanger (2023) show that the surge began as pandemic reallocation. AI could have extended it. The size of startups backed by VCs appears to have continued falling through 2025 (Carta, 2026). These figures do not distinguish the two mechanisms, and the model does not determine the distribution of firm sizes.

A.4 Construction of the measurement exhibits

The employment exhibits use BLS OEWS counts for the occupations and industries in Table 1. Specialist-industry shares divide employment in that industry by total employment in the occupation, including government. We combine the 2015–2020 MB3 research estimates with official estimates from 2021 onward, which use the same method (U.S. Bureau of Labor Statistics, 2021a,b, 2025). The older series shown in grey in Figure 2 uses a different method and is not directly comparable in levels. The long-run software series combines developers and quality-assurance analysts and testers; Table 1 and the 2022–2025 figures in the text use developers alone.

Each May estimate pools six semiannual panels over three years, so post-2022 changes enter gradually; five of the May 2025 panels postdate ChatGPT's release (U.S. Bureau of Labor Statistics, 2026). The estimates are subject to sampling error. The available data do not provide the variances and covariances needed to calculate standard errors for these shares.

Table A1. Share of each occupation employed in its main specialist industry, May 2015 to May 2025 (percent)

Function	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**2025**	**Trend**	**Change**	**Gap**
	MB3 research estimates						Official (MB3)					2015–22	2022–25	2025
						percent						pp/year	pp	
Legal	63.1	62.5	63.0	62.2	62.4	62.5	61.3	61.4	60.9	59.6	59.4	−0.23	−2.0	−1.4
Accounting	25.5	25.0	24.2	24.1	24.2	25.3	24.8	24.2	24.3	24.0	22.8	−0.08	−1.4	−1.4
Software[b]	36.2	36.5	36.5	35.7	34.7	34.3	33.4	34.2	32.7	31.4	30.5	−0.44	−3.6	−1.7
Management analysis	26.7	27.4	27.9	27.9	26.1	25.2	24.9	26.6	26.3	25.7	24.9	−0.27	−1.7	+0.1
Marketing	5.3	4.7	4.6	4.8	4.5	4.9	5.0	4.9	4.8	4.2	4.6	−0.02	−0.3	−0.2
Design	13.3	13.7	13.4	12.7	12.2	11.9	11.1	11.1	11.5	11.8	11.9	−0.40	+0.8	+2.0

Notes: For each function, the number of people in the occupation who work in the main specialist industry of Table 1, divided by the number in the occupation in the whole country, in all industries and in government, in percent. For 2015–2020 we use the MB3 research estimates that BLS computed with the method it adopted in May 2021, so all eleven years use one method. "Trend" is the slope of a straight line fitted to 2015–2022. "Change" is 2025 minus 2022. "Gap" is the 2025 value minus the trend line extended to 2025. BLS builds each May estimate from six semiannual panels collected over the previous three years, so a change that starts in late 2022 enters the numbers gradually; the May 2025 estimate still includes the November 2022 panel, so five of its six panels postdate ChatGPT's release. The estimates are subject to sampling error; standard errors for the shares are not reported. See Appendix A.4.

[b] Software adds software developers and software quality-assurance analysts and testers, because BLS counted them as one occupation in 2019 and 2020. Developers alone give 34.3 in 2022 and 30.5 in 2025.

Sources: Author's calculations from the BLS Occupational Employment and Wage Statistics: MB3 research estimates for May 2015 to May 2020 and national industry-specific estimates for May 2021 to May 2025 (U.S. Bureau of Labor Statistics, 2021b, 2025).

Table A2. Employment growth of each occupation in its main specialist industry and at other employers (percent per year)

Function	2015–2019 Specialist	2015–2019 Other	2019–2022 Specialist	2019–2022 Other	2022–2025 Specialist	2022–2025 Other	Employment 2025 (thousands) Specialist	Employment 2025 (thousands) Other
Legal	1.8	2.6	1.3	2.7	1.0	4.0	448	306
Accounting	0.2	1.8	3.3	3.3	-0.9	1.7	330	1,119
Software	4.4	6.1	6.9	7.8	-1.1	4.5	573	1,302
Management analysis	4.0	4.8	5.0	4.1	1.2	4.4	223	675
Marketing	2.9	7.3	6.8	4.3	1.9	4.2	41	859
Design	-1.6	0.9	-5.3	-1.8	0.0	-2.6	23	174

Notes: "Specialist" is the number of people in the occupation who work in the main specialist industry of Table 1. "Other" is the number who work for any other employer, including government. Growth rates are compound annual rates between May estimates. For 2015–2019 we use the MB3 research estimates; for 2019–2022 and 2022–2025 we use the 2019 research estimate and the official estimates from 2021 on. Software includes quality-assurance analysts and testers throughout. Sources: as in Table A1.

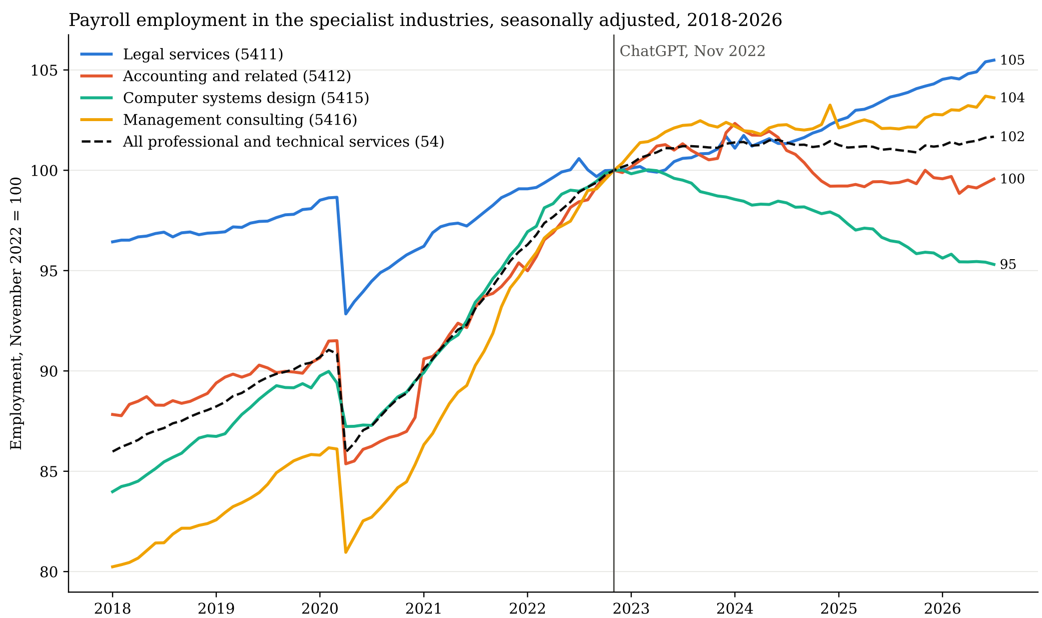

Figure A2. Payroll employment in four specialist industries and in all professional and technical services, monthly, seasonally adjusted, January 2018 to July 2026, with November 2022 set to 100. Source: BLS Current Employment Statistics, downloaded from FRED.

Figure A2 shows the payroll series behind the introduction. The slowdown partly reflects the end of the pandemic hiring boom: computer-services firms were 2 percent above their 2016–2019 trend in November 2022 and are 15 percent below it now, while law firms, which had no boom, are 2 percent above theirs.

Table A3 sums intermediate purchases of professional and business services by user sector from the BEA 2017 use table (before redefinitions, sector level, 2023 revision) (U.S. Bureau of Economic Analysis, 2023). These pre-AI figures measure gross intermediate use at producers' prices, not value added, and cover a broader commodity than the six functions in Table 1.

Table A3. Who purchases professional and business services

User sector	**Purchases ($ billions)**	**Share (percent)**
Professional and business services	581	18.5
Finance, insurance, and real estate	532	17.0
Wholesale and retail trade	510	16.3
Education, health, and social assistance	282	9.0
Government	273	8.7
Manufacturing	270	8.6
Information	225	7.2
Other user sectors	461	14.7
All user sectors	3,134	100.0

Notes: Intermediate uses of the commodity "Professional and business services" at producers' prices, rounded to the nearest billion. Gross purchases, not value added; the commodity is broader than the six functions in Table 1.
Source: Author's grouping of the BEA 2017 use table before redefinitions, sector level, 2023 revision (U.S. Bureau of Economic Analysis, 2023).

References

Abbott, Andrew. 1988. *The System of Professions: An Essay on the Division of Expert Labor*. Chicago: University of Chicago Press.

Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62(5): 777–795.

Association of Corporate Counsel and Everlaw. 2025. "Generative AI's Growing Strategic Value for Corporate Law Departments—Survey Results." ACC/Everlaw Survey Report, October 14.

Atalay, Enghin, Phai Phongthiengtham, Sebastian Sotelo, and Daniel Tannenbaum. 2020. "The Evolution of Work in the United States." *American Economic Journal: Applied Economics* 12(2): 1–34.

Bresnahan, Timothy F., and Shane Greenstein. 1999. "Technological Competition and the Structure of the Computer Industry." *Journal of Industrial Economics* 47(1): 1–40.

Chen, Ruiqi. 2022. "Microsoft, Activision Tap Simpson Thacher, Skadden for Game Deal." *Bloomberg Law*, January 18.

U.S. Bureau of Economic Analysis. 2023. "2017 Make–Use Framework: Use Table Before Redefinitions, Sector Level." 2023 revision.

U.S. Bureau of Labor Statistics. 2021a. "Occupational Employment and Wage Statistics: MB3 Research Estimates, May 2015–May 2020." https://www.bls.gov/oes/oes-mb3-methods.htm.

U.S. Bureau of Labor Statistics. 2021b. "Occupational Employment and Wage Statistics: May 2021 National and National Industry-Specific Occupational Employment and Wage Estimates."

U.S. Bureau of Labor Statistics. 2025. "May 2025 National and National Industry-Specific Occupational Employment and Wage Estimates."

U.S. Bureau of Labor Statistics. 2026. "Occupational Employment and Wage Statistics: Technical Notes for May 2025 Estimates." https://www.bls.gov/oes/current/oes_tec.htm.

Carta. 2026. "State of Startup Compensation: H2 2025." Carta Data, May 4.

U.S. Census Bureau. 2025. *2023 Business Dynamics Statistics Data Tables*. September 24.

U.S. Census Bureau. 2026. *Business Formation Statistics*. Monthly data through June 2026; annual averages use complete years through 2025.

Chin, Caroline, and Alex Martin Richmond. 2026. "Work at the Frontier: How AI Is Expanding What People Do at Work." OpenAI Economic Research, July 27.

Coase, Ronald H. 1937. "The Nature of the Firm." *Economica* 4(16): 386–405.

Crémer, Jacques. 1993. "Corporate Culture and Shared Knowledge." *Industrial and Corporate Change* 2(3): 351–386.

Crémer, Jacques, Luis Garicano, and Andrea Prat. 2007. "Language and the Theory of the Firm." *Quarterly Journal of Economics* 122(1): 373–407.

Decker, Ryan A., and John Haltiwanger. 2023. "Surging Business Formation in the Pandemic: Causes and Consequences?" *Brookings Papers on Economic Activity*, Fall: 249–302.

Dell'Acqua, Fabrizio, Charles Ayoubi, Hila Lifshitz, Raffaella Sadun, Ethan Mollick, Lilach Mollick, Yi Han, Jeff Goldman, Hari Nair, Stew Taub, and Karim R. Lakhani. 2026. "The Cybernetic Teammate: A Field Experiment on Generative AI and Teamwork." *Organization Science* 37(4): 1217–1242.

Demirci, Ozge, Jonas Hannane, and Xinrong Zhu. 2025. "Who Is AI Replacing? The Impact of Generative AI on Online Freelancing Platforms." *Management Science* 71(10): 8097–8108.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108(5): 874–904.

Garicano, Luis, and Thomas N. Hubbard. 2007. "Managerial Leverage Is Limited by the Extent of the Market: Hierarchies, Specialization, and the Utilization of Lawyers' Human Capital." *Journal of Law and Economics* 50(1): 1–43.

Garicano, Luis, and Thomas N. Hubbard. 2009. "Specialization, Firms, and Markets: The Division of Labor Within and Between Law Firms." *Journal of Law, Economics, and Organization* 25(2): 339–371.

Garicano, Luis, and Luis Rayo. 2025. "Training in the Age of AI: A Theory of Career Viability." CEPR Discussion Paper No. 20634. CEPR Press, Paris and London. Revised March 2, 2026.

Garicano, Luis, Jin Li, and Yanhui Wu. 2026. "Weak Bundle, Strong Bundle: How AI Redraws Job Boundaries." CEPR Discussion Paper No. 21453. CEPR Press, Paris and London.

Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94(4): 691–719.

Hadfield, Gillian K., and Andrew Koh. 2025. "An Economy of AI Agents." In *The Economics of Transformative AI*, edited by Ajay K. Agrawal, Erik Brynjolfsson, and Anton Korinek, chap. 5. Chicago: University of Chicago Press.

Hoffmann, Jordan, Sebastian Borgeaud, Arthur Mensch, et al. 2022. "Training Compute-Optimal Large Language Models." *Advances in Neural Information Processing Systems* 35: 30016–30030.

Hui, Xiang, Oren Reshef, and Luofeng Zhou. 2024. "The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market." *Organization Science* 35(6): 1977–1989.

Ide, Enrique, and Eduard Talamàs. 2025. "Artificial Intelligence in the Knowledge Economy." *Journal of Political Economy* 133(12): 3762–3800.

Knight, Frank H. 1921. *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin.

Muennighoff, Niklas, Alexander M. Rush, Boaz Barak, et al. 2023. "Scaling Data-Constrained Language Models." *Advances in Neural Information Processing Systems* 36: 50358–50376.

Retkowsky, Jana, Ella Hafermalz, and Marleen Huysman. 2026. "Harnessing a 'Spirited Technology': How Working with Generative AI Collapses the Creative Process." *Academy of Management Discoveries*. Advance online publication, April 20. `https://doi.org/10.5465/amd.2024.0280`.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98(5, Part 2): S71–S102.

Rosen, Sherwin. 1983. "Specialization and Human Capital." *Journal of Labor Economics* 1(1): 43–49.

Shahidi, Peyman, Gili Rusak, Benjamin S. Manning, Andrey Fradkin, and John J. Horton. 2025. "The Coasean Singularity? Demand, Supply, and Market Design with AI Agents." NBER Working Paper No. 34468.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan and T. Cadell.

Stigler, George J. 1951. "The Division of Labor Is Limited by the Extent of the Market." *Journal of Political Economy* 59(3): 185–193.

Thomson Reuters Institute and the Center on Ethics and the Legal Profession at Georgetown Law. 2026. "2026 Report on the State of the US Legal Market." January.