International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

The Economics of Artificial Intelligence in U.S. Healthcare: Productivity Gains, Value-Chain Reconfiguration, and the Future of Medical Labor

Mr. Saavya Sharma[1], Mr. Shreyansh G[2]

Abstract
Artificial intelligence (AI) offers one of the more credible near-term routes to higher productivity in the U.S. healthcare sector, an industry long marked by high costs, negative labor-productivity growth, and persistent quality shortfalls. This paper synthesizes and extends two recent contributions from the NBER volume *The Economics of Artificial Intelligence: Health Care Challenges*. Using detailed domain-level estimates, Sahni and coauthors show that wider adoption of today's machine-learning and natural-language-processing tools could generate $200–360 billion in annual net savings (2019 dollars) within five years—roughly 5 to 10 percent of total spending—without sacrificing quality or access. At the same time, a task-based reading of physician work, following the Autor–Levy–Murnane framework and developed by Dranove and Garthwaite, indicates that AI will complement some specialties while substituting for others, particularly those whose core tasks are closer to pattern recognition on structured data. Historical patterns of third-party intervention in medical decision-making, together with well-documented organizational and industry-level barriers, imply that diffusion will be uneven and that the distribution of economic rents along the value chain is far from predetermined. The paper expands the original analyses with additional discussion of general-equilibrium effects, specialty-level exposure, implementation cost recovery, and policy design. Realizing the productivity gains will require deliberate attention to incentives, data infrastructure, talent, and the political economy of professionalstatus.

Keywords: artificial intelligence, healthcare productivity, medical labor, value chain, automation, cost savings, physician agency

1. Introduction

It is often said, only half in jest, that the most expensive piece of medical technology remains the physician's pen. The aphorism captures a deeper truth: physicians still sit at the center of the healthcare value chain, diagnosing, recommending, and directing the application of other inputs. Victor Fuchs labeled them the "captain of the team" half a century ago, and the basic structure has not changed much even as the team has grown larger and far more expensive (Fuchs 1974). Patients continue to place high trust in physicians relative to other occupations, yet that trust coexists with well-documented fallibility—both under-treatment of the seriously ill and over-treatment of the largely healthy (Mullainathan and Obermeyer 2021). The result is the familiar and undesirable combination of high spending and preventable harm. Against this background, artificial intelligence has been held out as a possible productivity lever.

Machine learning can learn patterns from large data sets without being programmed with explicit rules; natural-language processing can turn unstructured clinical notes and patient communications into structured, actionable information. In principle these tools can simplify repetitive administrative processes, create new ones that were previously infeasible, and supply clinicians with better predictions and recommendations at the moment of decision. The question is no longer whether AI can contribute something useful, but how large the contribution might be, how quickly it can diffuse, and who will capture the resulting economic value.

Two recent papers in the NBER conference volume on the economics of AI in healthcare provide complementary answers. Sahni, Stein, Zemmel, and Cutler (2024) assemble a bottom-up estimate of potential cost savings from AI adoption across hospitals, physician groups, and private payers. They conclude that technologies already available today could, if scaled within five years, reduce U.S. healthcare spending by 5 to 10 percent, or $200–360 billion annually in 2019 dollars, while preserving quality and access. Dranove and Garthwaite (2022) take a longer historical and labor-economics view. They treat AI as the latest stage in a decades-long sequence of third-party attempts to improve medical decision-making, and they apply a task-based framework to ask which physician specialties are most exposed to substitution versus complementarity.

This paper synthesizes those two contributions, expands the economic analysis, and adds several layers of discussion that the original papers could only touch lightly. Section 2 reviews the historical attempts to influence physician decision-making and situates AI within that lineage. Section 3 presents the savings estimates in greater detail, including methodology, domain-level breakdowns, and illustrative case studies. Section 4 develops the task-based account of physician work and explores specialty-level implications. Section 5 examines organizational and industry-level barriers to adoption.

Section 6 considers market trends, general-equilibrium effects, and the recovery of implementation costs. Section 7 turns to policy design. A short conclusion follows.

3A word on tone and evidence. The authors of the Sahni et al. paper include both academics and consultants with extensive field experience; many of the quantitative estimates therefore rest on bserved implementations and expert judgment rather than randomized trials. Dranove and Garthwaite are frank that their analysis of specialty exposure is partly conjectural. I follow their lead in being explicit about where the evidence is thinner. The goal is not to claim precision beyond what the data support, but to map the main economic mechanisms and the open empirical questions.

One further preliminary remark is worth making. Healthcare is unusual among large sectors in the degree to which the central decision-maker—the physician—is also a residual claimant on the volume and mix of services. In manufacturing or logistics, the firm that installs robots typically captures most of the productivity gain. In healthcare the physician’s dual role as agent and residual claimant creates both a potential conflict of interest and a political obstacle to technologies that threaten that role. Any serious economic analysis of AI in this sector has to keep that dual role in view.

The remainder of the paper proceeds as follows. After the historical review in Section 2, Section 3 unpacks the savings calculations in more detail than the original summary tables allow, including a discussion of how net savings are derived from gross and of the sensitivity of the aggregates to alternative assumptions about public payers and other sites of care. Section 4 develops the task-based framework at greater length and offers a more granular, if still provisional, ranking of specialty exposure. Section 5 expands the treatment of barriers, distinguishing those that individual organizations can address from those that require

collective action. Section 6 turns to dynamics—capital flows, cost recovery, induced demand, and labor-market adjustment. Section 7 draws out policy implications.
Section 8 concludes.

2. Historical Context: Third-Party Intervention and the Physician as Agent

Arrow’s classic 1963 essay remains the natural starting point. Because patients are imperfectly informed and the stakes are high, they rely on physicians as learned agents. Professional training, ethical norms, and peer pressure are supposed to keep agency problems in check. In practice the check has been incomplete. Fee-for-service payment has long been suspected of inducing excess volume; geographic practice variations have been documented for decades; and the Institute of Medicine’s 2000 report estimated that tens of thousands of preventable deaths occur each year from medical error (Wennberg and Gittelsohn 1973; Cutler et al. 2019; Donaldson et al. 2000).

Attempts to rein in these problems pre-date modern computing. Second-surgical-opinion programs appeared in the 1950s. The Social Security Amendments of 1972 created Professional Standards Review Organizations (PSROs)—local physician panels charged with developing “objective” standards and reviewing claims. By the late 1970s a congressional subcommittee claimed that more than two million unnecessary surgeries were being performed annually. Congress responded by adding second-opinion requirements for Medicare. The PSROs, however, lacked strong enforcement powers and their guidelines were not widely disseminated; the successor Peer Review Organizations of the 41980s fared little better (Gray and Field 1989).

Commercial insurers proved more aggressive. Utilization-review (UR) firms such as InterQual developed explicit intensity-of-service and severity-of-illness criteria, implemented them with early computer algorithms, and refused payment when criteria were not met. Physicians and patients pushed back hard. Surveys in the late 1990s found that a majority of physicians reported serious problems with external review. Complaints centered on opacity, the omission of soft information that only the treating clinician possessed, and the sheer time cost of compliance. Politicians noticed; several states enacted laws exposing insurers to malpractice liability for UR decisions, and federal “patient bill of rights” proposals sought to curb overrides of physician judgment. By the early 2000s most insurers had retreated from punitive UR toward purely informational clinical decision support (Wickizer and Lessler 2002; Dranove 1993).

The retreat is itself informative. Research on the cost-saving effects of utilization review produced mixed results; some studies found modest reductions in hospital use, others found little net effect once the administrative costs of the review process were counted. More important for the politics of the episode was the perception among physicians that their professional judgment was being second-guessed by remote nurses and opaque algorithms, and the perception among patients that needed care was being delayed or denied. Those perceptions fueled a regulatory and legislative response that ultimately limited the most aggressive forms of review. The episode illustrates a recurring pattern: tools that appear to transfer decision rights away from the treating clinician generate organized resistance, even when the tools have a plausible efficiency rationale. AI systems that ignore this history risk repeating it.

At the same time, the informational turn in clinical decision support created a platform on which later AI tools could build. Once hospitals and practices had electronic health records and once payers and providers had grown accustomed to receiving protocol-based suggestions, the marginal step to more sophisticated predictive models became smaller. In that sense the mixed legacy of utilization review and evidence-based medicine is both a warning about political constraints and a partial foundation for the next wave of technology.

The subsequent evidence-based-medicine movement—Patient Outcome Research Teams, treatment protocols embedded in electronic health records, and the broader clinical-decision-support infrastructure—has remained largely advisory. Most physicians report that the positive aspects of protocols outweigh the negatives, though a sizable minority still feel their clinical autonomy is constrained (Deloitte 2016 Survey of U.S. Physicians). AI is the natural next step: algorithms that can ingest far more granular data, generate finer-grained recommendations, and, in some domains, perform the diagnostic reading itself. The historical pattern suggests both opportunity and friction. Tools that improve decision quality while preserving a meaningful role for the clinician have diffused more readily than tools that appear to demote the physician from captain to technician.
5It is useful to distinguish two mechanisms through which third-party tools can operate. The first is complementary: the tool supplies information or a recommended action that the physician is free to accept, modify, or reject. The second is substitutive: the tool's output becomes binding, or the cost of ignoring it (in the form of denied payment, liability exposure, or employer discipline) becomes high enough that the physician's residual discretion shrinks. PSROs and early evidence-based protocols were closer to the complementary end of the spectrum; commercial utilization review in its aggressive phase was closer to the substitutive end and provoked the stronger backlash. Modern AI systems can be designed either way. The design choice will influence both the speed of adoption and the ultimate distribution of authority and rents along the value chain.
A related point concerns the information that algorithms historically omitted. Utilization-review criteria could capture measurable clinical parameters—laboratory values, vital signs, documented diagnoses—but they struggled with the soft data that a treating physician observes: how the patient describes pain, whether the home environment will support a particular regimen, the patient's track record of adherence, the nuances of preference under uncertainty. Physicians repeatedly cited these omissions as a reason to resist algorithmic overrides. Contemporary machine-learning systems can, in principle, ingest richer data, including free-text notes and even some forms of patient-generated information. Whether they do so in practice, and whether clinicians trust the resulting recommendations, remains an open empirical question and a central design challenge for vendors and health systems alike.

3. The Near-Term Savings Opportunity

3.1 Scope and Method

Sahni et al. (2024) deliberately restrict attention to technologies that already exist and to adoption that could plausibly reach scale within five years. They exclude longer-horizon possibilities such as digital twins or fully generative clinical systems, and they set aside AI applications further up the value chain (drug discovery, medical-device design, medical education). The focus is therefore on operational and clinical processes inside hospitals, physician groups, and private payers—the three stakeholder groups that together account for roughly 80 percent of industry revenue.
The estimation strategy is bottom-up. Start with 2019 National Health Expenditure totals for each stakeholder group. Subtract earnings before interest and taxes (drawn from proprietary value-pool data) to isolate total costs. For hospitals and physician groups, partition costs into administrative, labor-related medical, and non-labor medical components; for private payers the partition is simply administrative versus medical. Map each AI domain onto the relevant cost category, apply a net savings rate (gross savings minus ongoing maintenance costs for labor and technology), and multiply through. One-time

implementation costs—typically 1.0 to 1.5 times annual net savings—are acknowledged but excluded from the run-rate figures.

6Two conceptual distinctions run through the analysis. First, some use cases simplify existing processes (for example, automated claims adjudication); others create processes that were previously impractical (for example, real-time deterioration prediction at scale). Second, savings can fall on administrative costs or on medical costs; the latter often operate through better targeting of care rather than pure price or wage reduction. The authors emphasize that the estimates assume no sacrifice of quality or access; in several domains the non-financial gains (safety, experience, clinician time) are part of what makes the business case viable under a "total mission value" framing.

3.2 Domain-Level Results

Hospitals face total costs of about $1,096 billion (2019). The projected annual net savings range is $60–120 billion, or roughly 5–11 percent of costs. Clinical operations (operating-room block optimization, capacity and workflow management, diagnostics logistics, supply-chain forecasting, clinical workforce scheduling) and quality-and-safety applications (early warning for deterioration, adverse-event detection, readmission risk) are the primary drivers. Approximately 40 percent of the savings come from administrative costs and 60 percent from medical costs; the split between simplifying existing processes and creating new ones is nearly even.

Physician groups, with total costs around $711 billion, could realize $20–60 billion (3–8 percent). Clinical operations again dominate—missed-visit prediction, outpatient capacity management, care-team deployment—supplemented by continuity-of-care and value-based-care use cases such as targeting patients at high risk of unplanned admission. Savings split roughly evenly between administrative and medical costs.

Private payers show the largest absolute opportunity: $80–110 billion on a cost base of $1,135 billion (7–10 percent). Claims management (auto-adjudication, prior-authorization prediction, fraud-waste-abuse detection), health-care management (tailored outreach, avoidable-readmission reduction), and provider-relationship management (network design, directory accuracy, value-based contracting support) account for most of the gains. Here medical-cost reductions predominate, roughly 80 percent of the total.

Scaling to public payers (assumed to have costs about 45 percent of private payers and a somewhat smaller relative opportunity) and other sites of care (dentists, home health, etc., assumed to have costs about 115 percent of physician groups but only half the relative savings rate) produces the industry-wide total of $200–360 billion, or 5–10 percent of national health spending. Administrative costs overall could fall 7–14 percent ($65–135 billion); medical costs 5–8 percent ($130–235 billion).

Stakeholder	**Total Costs ($bn, 2019)**	**Net Savings ($bn)**	**% of Costs**	**Share Admin**
Hospitals	1,096	60–120	5–11%	~40%
Physician groups	711	20–60	3–8%	~50%
7Private payers	1,135	80–110	7–10%	~20%
Public payers	511	30–40	5–7%	~20%
Other sites of care	817	10–30	1–4%	~50%
Total		**200–360**	**5–10%**	**~35%**

Table 1. Estimated annual net savings opportunity from AI adoption within five years using current techn-

ologies (2019 dollars). Source: adapted from Sahni et al. (2024), Table 2.1.

To put the aggregates in perspective, a 5–10 percent reduction in national health spending is large relative to the productivity gains the sector has achieved through earlier waves of organizational and technological change. Administrative-cost shares have been stubbornly high for decades; labor-productivity growth in healthcare has lagged most other service industries. If even the lower end of the estimated range were realized and sustained, the cumulative fiscal and private-sector impact over a decade would be measured in the trillions. That arithmetic is what makes the adoption barriers, and the policy levers that might lower them, first-order economic questions rather than secondary implementation details.

Domain cluster Primary stakeholders Main cost category Illustrative use cases

Clinical operations Hospitals, physician groups Medical + admin OR optimization, capacity, staffing, supply chain Quality & safety Hospitals, physician groups Medical (labor) Deterioration prediction, adverse-event detection, readmission risk Claims management Private (and public) payers Admin + medical Auto-adjudication, prior auth, FWA detection Care / health management Payers, value-based Providers Medical Tailored outreach, avoidable admissions, behavioral health matching Corporate & member services

All Administrative Call routing, virtual agents, directory accuracy

Table 2. Illustrative mapping of domain clusters to stakeholders, cost categories, and use cases. Adapted from the domain
frameworks in Sahni et al. (2024).

3.3 Illustrative Case Studies

Several concrete implementations help make the mechanisms tangible. During the COVID-19 surge a multistate hospital system faced sharply higher call volumes and long wait times. Using atural-language processing it fielded a virtual agent on its app and website that answered common questions and routed residual queries to the right queue. Call volume fell by nearly 30 percent; patient experience scores improved; and staff were redeployed to more complex, less scriptable interactions. A large regional hospital discovered that its operating rooms appeared fully blocked yet actual utilization hovered around 60 percent. Historical block schedules did not adjust for surgeon-specific demand, slot lengths were poorly calibrated to actual case times, and release of unused time was manual. An optimization algorithm that ingested utilization history, forecast demand by specialty and surgeon, and respected equipment and staffing constraints raised open capacity by about 30 percent. 8Sustaining the gain required assigning a data scientist ongoing responsibility and keeping surgeons in the loop so that the recommendations retained legitimacy.

On the physician-group side, one organization in a value-based contract for a chronic condition observed that roughly 10 percent of its patients were admitted each month. A risk model trained on electronic health records, laboratory results, demographics, and admission-discharge-transfer feeds identified patients at elevated risk of unplanned admission and the variables that drove that risk. Care teams could then prioritize outreach; early prototypes suggested several percentage-point reductions in inpatient spending.

Among private payers, a fraud-waste-abuse classification model trained on multi-year claims data produced a prioritized list of providers for investigation and contributed to a roughly 50-basis-point reduction in medical costs. Separately, a readmission-prediction model combined with redesigned outreach raised the share of high-risk members who connected with care managers by about 70 percent, increased 30-day follow-up visits by 40 percent, and cut all-cause readmissions in the target cohort by approximately 55 percent.

These examples share several features. They address processes that are either highly repetitive or that involve complex but learnable decision trees. They combine prediction with redesigned workflows rather than treating the algorithm as a stand-alone product. And they required ongoing human oversight—both technical maintenance and clinical or operational buy-in—to sustain performance. It is also worth noting what the case studies do not show. They do not demonstrate that the same algorithms would perform equally well in organizations with weaker data infrastructure, less engaged clinical leadership, or different patient populations. External validity remains a concern. Nor do they isolate the pure contribution of the algorithm from the contribution of the accompanying workflow redesign. In several instances the organizational change appears to have been at least as important as the predictive model itself. This observation has implications for how we should interpret the aggregate savings estimates: those estimates implicitly assume that the complementary organizational investments will be made at scale, which is a strong assumption.

A further distinction that runs through the domain analysis is the difference between administrative and medical cost savings. Administrative processes—claims adjudication, prior authorization, call routing, provider directory management—are often more standardized and therefore more immediately amenable to automation. Medical-cost savings typically require changing clinical behavior or targeting interventions more precisely; they are harder to achieve and easier to reverse if clinicians disengage. The fact that private-payer savings are skewed toward medical costs while hospital and physician-group savings are more balanced therefore signals both larger opportunity and greater implementation difficulty for the payer use cases that rely on influencing downstream clinical decisions.

3.4 Caveats on the Savings Estimates

9A few limitations deserve emphasis. First, the net savings rates are grounded in observed implementations and expert judgment rather than a large body of experimental evidence; the confidence intervals are therefore wider than the reported ranges might suggest. Second, the estimates assume full-scale adoption. In reality diffusion will be incomplete and staggered, so realized savings over any five-year window will be lower. Third, one-time implementation costs are excluded from the run-rate figures; organizations with limited capital or change-management capacity may find the upfront outlay prohibitive even when the steady-state return is attractive. Fourth, the analysis holds quality and access constant by assumption; if AI tools are deployed carelessly, quality could deteriorate in ways that offset financial gains. Finally, general-equilibrium responses—induced demand, wage adjustments, entry of new providers—are outside the partial-equilibrium frame and could either amplify or erode the net social saving.

Sensitivity to the public-payer and other-sites assumptions is also worth noting. The paper scales from private-payer estimates by assuming public payers have costs equal to about 45 percent of private payers and a relative savings opportunity about three-quarters as large, reflecting a narrower set of functions. Other sites of care are scaled from physician-group estimates at 115 percent of costs and half the relative savings rate, reflecting differences in acuity and clinical intensity. Alternative assumptions—higher relative opportunity for public payers if Medicare Advantage and Medicaid managed care converge toward private-payer practices, or lower opportunity for other sites if their processes are less algorithm-friendly—would move the industry total within a band that is still measured in the low-to-mid hundreds of billions.

The qualitative conclusion that the opportunity is large relative to historical productivity gains is robust to these variations.

Another methodological choice is the focus on run-rate net savings after maintenance costs but before one-time implementation outlays. In a discounted-cash-flow sense the correct metric is the present value of the net stream, which depends on the discount rate, the time path of adoption, and the durability of the savings. Organizations with high costs of capital or short managerial horizons will apply a higher effective discount rate and will therefore require larger or faster savings to justify investment. This is one reason that payment models with multi-year shared-savings horizons, or public programs that can take a longer view, may be better positioned to finance the transition than purely commercial entities under quarterly earnings pressure.

4. AI, the Value Chain, and the Task Content of Physician Work

4.1 The Stylized Value Chain

A simplified sequence helps fix ideas. A patient presents with symptoms or for preventive care. The clinician gathers information—history, examination, and ordered tests. A diagnosis is reached. A treatment plan is recommended and, if accepted, delivered. Throughout, the physician has traditionally acted as both information aggregator and decision-maker, capturing a substantial share of the value created. Average physician incomes near $350,000 (with surgeons higher and primary-care physicians 10lower) place them among the top earners in the economy (Gottlieb et al. 2020).

Other actors—hospitals, device and drug firms, insurers—also create and capture value. Patients receive the residual surplus in the form of health benefits net of premiums and out-of-pocket costs. Anything that changes the relative productivity or scarcity of a particular input can shift the division of rents. AI is precisely such a force.

4.2 Routine versus Non-Routine Tasks

Autor, Levy, and Murnane (2003) observed that early computerization substituted most readily for routine tasks—those that follow explicit, programmable rules—while raising the relative demand for non-routine analytic and interactive tasks that computers could not then perform. Subsequent waves of automation, including industrial robots, continued to displace routine work (Acemoglu and Restrepo 2020). AI pushes the frontier further: pattern recognition on high-dimensional data, once the province of human experts, is increasingly automatable.

In medicine the distinction maps onto the stages of the value chain. When the relevant information is largely physical and structured—radiographic images, histopathology slides, laboratory panels—and the mapping from data to diagnosis can be learned from large labeled sets, AI can perform the reading at least as accurately as the average specialist and often more consistently. Radiology and pathology are the clearest current examples. In those domains the specialist's residual role may shrink to oversight, quality control, or the handling of edge cases. Value can migrate toward the owners of the algorithms, the platforms that integrate them, or the lower-cost clinicians who act on the outputs. When critical information must still be elicited through conversation—affect, adherence barriers, social context, idiosyncratic preferences—AI is more likely to complement the clinician who is skilled at gathering and interpreting that soft information. The same logic applies to treatment planning that requires negotiation with the patient or coordination across multiple chronic conditions. Specialties whose comparative advantage lies in these non-routine interactive and contextual tasks face less immediate substitution pressure and may see their productivity, and possibly their earnings, rise if AI handles the more routine cognitive load.

4.3 Specialty-Level Exposure: A Tentative Ranking

A precise ranking would require detailed task inventories and measures of AI performance by task; those data are still incomplete. Still, a qualitative ordering is possible. At higher substitution risk sit radiology, pathology, and certain domains of dermatology and ophthalmology where image-based diagnosis predominates. Anesthesiology has already seen algorithmic support for monitoring; further automation of routine cases is plausible. At intermediate risk are specialties that mix pattern recognition with procedural skill or contextual judgment—cardiology (especially imaging and electrophysiology), oncology (treatment selection from genomic and clinical data), and parts of orthopedics. At lower near-term risk are primary care, geriatrics, psychiatry, and complex multi-morbidity management, where the soft-information and relationship-intensive components remain large. 11Even within high-risk specialties the picture is not pure displacement. AI that is more accurate than the average reader can raise the productivity of the remaining human experts who handle the difficult cases or who integrate imaging findings with the rest of the clinical picture. The net employment and wage effects will depend on the elasticity of demand for the specialty's services, the ease of entry, and the degree to which complementary non-routine tasks expand. Historical analogies—the effect of spreadsheets on bookkeepers, or of GPS on taxi drivers—suggest that average wages in automatable occupations can stagnate or fall even while residual specialists capture rents.

Radiology offers a concrete illustration of the forces at work. Multiple studies have shown that convolutional neural networks can match or exceed average radiologist performance on specific detection tasks (for example, certain findings on chest radiographs or mammograms). If such systems are deployed at scale, the volume of studies that require a full specialist read may fall, or the time per study may decline. In a fee-for-service environment the immediate effect could be pressure on radiologist incomes; in an employment or risk-bearing environment the effect could be a reallocation of radiologist time toward complex cases, interventional procedures, or multidisciplinary conferences. The long-run supply response will depend on how training programs and certification requirements adapt. Similar dynamics are already visible in pathology, where whole-slide imaging and algorithmic support are beginning to change laboratory workflows.

Primary care sits at the opposite end of the spectrum for the near term. The work involves extensive soft-information gathering, longitudinal relationship management, coordination across specialists, and shared decision-making under uncertainty—all tasks that remain difficult to automate. AI can still raise primary-care productivity by handling documentation, surfacing relevant guidelines, predicting no-shows, or identifying patients who need outreach. In that complementary role the technology is more likely to expand the effective capacity of the existing workforce than to substitute for it. Whether the resulting gains accrue to primary-care physicians as higher earnings, to patients as improved access, or to payers as lower total cost of care will depend on payment design and on the organization of primary-care practices.

Between these poles lie mixed cases. Cardiology combines image-intensive diagnostics with complex clinical judgment and procedures; oncology increasingly relies on genomic and molecular data that algorithms can help interpret, yet treatment selection still involves preferences, comorbidities, and evolving evidence. In these specialties the division of labor between human and machine is likely to be contested and to evolve with the technology. Empirical work that tracks time use, task allocation, and relative wages within specialties as AI tools diffuse would be highly valuable for refining the predictions offered here.

4.4 Mid-Level Providers and the Redistribution of Tasks

If AI reliably handles a larger share of diagnostic and protocol-driven decisions, the marginal value of a fully trained physician in some settings may decline relative to the value of a nurse practitioner or physician assistant who can collect the necessary information and act on the algorithmic 2recommendation. Retail clinics and virtual-first primary-care models already experiment with this division of labor. Whether mid-level providers capture much of the resulting surplus depends on the elasticity of their labor supply and on professional and regulatory barriers that limit substitution. In the short run sticky wages and restricted entry can produce rents; in the longer run entry and scope-of-practice expansion are likely to erode them.

Hospitals that employ physicians and that bear financial risk under bundled or capitated contracts have stronger incentives to reoptimize the skill mix than do independent physician groups paid fee-for-service. This asymmetry helps explain why private payers and large health systems appear further along the adoption curve than smaller independent practices. One can push the task-based logic a step further by borrowing the language of levels of automation developed in other domains. The Society of Automotive Engineers defines six levels for driving automation, from no automation to full autonomy. Sahni et al. adapt a simplified version of that ladder to healthcare. Clinical decision-making in which the physician retains final authority while receiving algorithmic suggestions sits near level 1. Automated reading of radiology images with human quality control sits near level 2. Referral recommendations or claims adjudication that run with limited human intervention sit higher. The higher the level, the greater the potential productivity gain and the greater the threat to existing professional roles. Most current clinical deployments remain at the lower levels; the economic and political questions become sharper as systems move up the ladder. It is also important to recognize that productivity in the medical context is not solely a matter of speed or cost. Following the Autor–Levy–Murnane intuition, productivity can be thought of as the proximity of the reported diagnosis or treatment recommendation to the (unobserved) true optimum for the patient, taking both health outcomes and resource use into account. AI can raise that productivity by reducing both false positives and false negatives, by incorporating covariates that a busy clinician might overlook, and by enforcing consistency across cases. Whether the higher measured productivity translates into higher wages or employment for a given type of clinician depends on demand elasticities and on the degree of complementarity with remaining human tasks. In markets where demand for accurate diagnosis is relatively inelastic, labor that is complementary to AI may capture substantial rents; where demand is elastic or where entry is easy, the gains are more likely to be competed away or captured by patients and payers in the form of lower prices or premiums.

Finally, the possibility of biased algorithms deserves explicit mention. Obermeyer and coauthors have shown that a widely used commercial algorithm for predicting health needs exhibited racial bias because it used cost as a proxy for need, and costs themselves reflected unequal access. AI does not automatically eliminate the biases that have long characterized human medical decision-making; it can encode them, amplify them, or, with careful design and auditing, mitigate them. The distributional consequences of AI adoption therefore include not only the division of rents between capital and labor or between specialties, but also the potential for differential effects across patient groups. Any claim that AI will improve average productivity must be accompanied by attention to the variance and the 13equity of those improvements.

5. Barriers to Adoption

Despite the projected returns, healthcare has lagged other sectors in AI adoption (Cam, Chui, and Hall 2019). Financial services use sophisticated models for fraud, credit, and customer targeting; mining and retail have automated forecasting and logistics at scale. Healthcare remains earlier on the classic S-curve for most of the domains reviewed above. Both organizational ("within") and industry-level ("between" and "seismic") factors contribute.

5.1 Organizational Factors

Sahni et al. (2024) distill six requirements for successful deployment. First is a mission-led roadmap that links AI use cases to business objectives and to the broader mission, and that sequences implementation realistically. In healthcare the relevant objective function includes non-financial outcomes—quality, safety, patient and clinician experience, access—so traditional ROI calculations often undervalue projects. Organizations that frame the case in terms of "total mission value" find it easier to clear investment hurdles.

Second is talent. AI skills are scarce economy-wide; healthcare organizations face the additional difficulty that many are not located in talent hubs and cannot easily match technology-sector compensation. Third are the technology stack and tools—legacy electronic health records, limited interoperability, and the cost of creating clean training data. Fourth is data management itself: governance, quality, lineage, and privacy. Fifth is an agile delivery model that can iterate rather than follow multi-year waterfall plans. Sixth is change management: clinicians and staff must trust the tools enough to alter workflows, which in turn requires transparency, local adaptation, and visible respect for professional judgment.

Private payers, especially large national ones, have progressed furthest; hospitals are piloting andbeginning to scale in selected domains; most independent physician groups remain at an early stage unless they are employed by systems that push adoption from the center.

5.2 Industry-Level and Institutional Barriers

Even perfect internal execution would leave several external obstacles. Data heterogeneity and the absence of widely adopted interoperability standards raise the cost of assembling the large, clean training sets that modern models require. Patient confidence is not automatic; high-profile failures or perceived opacity can slow diffusion. Misaligned incentives under residual fee-for-service payment reduce the private return to cost-reducing innovation, particularly when the savings accrue to a different rganization from the one that bears the implementation cost. Industry fragmentation means that many of the highest-value use cases (continuity of care, network optimization) involve externalities across firms. Finally, regulatory and liability uncertainty—who is responsible when an AI recommendation 14contributes to a bad outcome?—creates option value to waiting.

The political economy of professional status adds another layer. Physicians have historically resisted tools that appear to demote them from autonomous professionals to technicians executing algorithmic instructions. Utilization review in the 1990s provoked organized opposition precisely because it threatened that status. AI systems that are framed as decision support, that keep the clinician in the loop for final authority, and that demonstrably reduce rather than increase cognitive burden are more likely to gain acceptance than systems that are perceived as replacing judgment.

Talent markets interact with these institutional barriers. Healthcare organizations compete for data scientists and machine-learning engineers against technology firms that offer higher compensation, equity upside, and more flexible work environments. Establishing satellite talent hubs in cities with deeper technical labor pools is one response, but it is available mainly to large systems and national payers.

Smaller hospitals and independent groups face a harder constraint and may remain dependent on vendors whose incentives are not perfectly aligned with local clinical priorities. The resulting dual structure—sophisticated AI deployment in well-resourced organizations, limited or purely off-the-shelf adoption elsewhere—could widen existing disparities in operational performance and patient experience.
Data governance and privacy rules, while essential for patient trust, also raise the cost of the large-scale data assembly that modern models require. Differing state laws, varying institutional review-board practices, and the residual effects of earlier data-blocking behaviors all contribute. The federal push toward greater interoperability has improved the technical feasibility of data exchange, yet the economic incentives to share data that might advantage a competitor remain weak.
Collective-action problems of this sort are classic candidates for public intervention or for industry consortia with credible enforcement mechanisms; neither has yet fully solved the problem.
A final organizational observation concerns the measurement of success. Many early AI projects in healthcare have been evaluated on narrow technical metrics—area under the curve, positive predictive value—without rigorous assessment of downstream clinical or financial impact. The gap between model performance in a retrospective data set and value creation in a live operational setting is often large. Organizations that treat AI as a technology procurement problem rather than as a sociotechnical change problem tend to under-invest in the workflow redesign, training, and continuous monitoring that convert predictions into better decisions. Closing that gap is as much a management challenge as a technical one.

6. Market Trends, Implementation Costs, and General-Equilibrium Considerations

6.1 Capital and Vendor Dynamics

15Venture and private-equity investment in healthcare AI has risen sharply in recent years. The influx of capital can accelerate product development and lower the fixed cost of adoption for providers who can purchase rather than build. At the same time, a crowded vendor landscape raises search and contracting costs, and many early products have struggled to demonstrate durable clinical or financial impact once deployed outside the development environment. Consolidation among vendors and the emergence of clearer performance benchmarks would reduce these frictions.

6.2 Recovering One-Time Costs

Because one-time implementation costs are often 1.0–1.5 times the annual net savings, the payback period is typically measured in years rather than months. Organizations with short managerial horizons, high discount rates, or binding capital constraints may forgo projects whose social return is positive. Shared-savings contracts, vendor financing, or public subsidies for data-infrastructure investments can improve the private calculus. Multi-year budgeting that explicitly capitalizes change-management and training costs, rather than treating them as operating expenses that must be absorbed in a single year, also helps.

6.3 Induced Demand and Volume Responses

Lower marginal cost of diagnosis or of certain treatments can induce additional volume. If AI makes imaging cheaper and faster, more imaging may be ordered; if risk prediction identifies more patients who could benefit from intensive care management, utilization of those services may rise. Whether the net effect on spending is still negative depends on the elasticity of demand and on whether the extra volume is high-value or low-value. In systems that remain largely fee-for-service the risk of low-value inducement is higher; under prospective payment or capitation the incentive to restrain low-value volume is stronger. This is one channel through which payment reform and AI adoption are complementary.

6.4 Labor-Market Adjustments

If AI reduces the demand for certain physician tasks, medical-school applications, residency choices, and ultimately specialty supply will adjust—but with long lags. In the interim, relative wages and vacancy rates can signal the shift. Allied-health occupations may expand more quickly if scope-of-practice rules permit. The net effect on total healthcare employment is ambiguous: automation of some tasks can raise productivity enough to expand output and employment in complementary tasks, as has occurred in other sectors. What is clearer is that the composition of the workforce, and the skill mix within it, will change.

Consider a simple comparative-static exercise. Suppose AI raises the productivity of diagnostic imaging interpretation by a factor of two while leaving the productivity of the subsequent treatment decision unchanged. If the demand for accurate diagnoses is relatively inelastic (patients and payers still want the diagnoses, just at lower cost), the total resources devoted to interpretation may fall, and the residual human interpreters who handle complex cases may earn higher wages. If, however, cheaper 16and faster interpretation induces a large increase in imaging volume—some of it low-value—the net spending reduction can be smaller than the partial-equilibrium calculation suggests, and employment in related services (technologists, follow-up clinics) may rise. The same logic applies, with different elasticities, to care-management outreach, prior-authorization automation, and operating-room scheduling. Partial-equilibrium savings estimates should therefore be read as upper bounds pending evidence on demand responses.

Another dynamic consideration is learning-by-doing and model drift. Clinical populations change, coding practices evolve, and the relationship between predictors and outcomes can shift. Models that are not continuously monitored and retrained will degrade. The ongoing maintenance cost that Sahni et al. subtract when moving from gross to net savings is therefore not a one-time accounting adjustment; it is a real and recurring operational requirement. Organizations that treat AI as a capital project that ends at go-live will eventually see performance erode. Those that build the corresponding operating model—data pipelines, monitoring dashboards, clinical feedback loops—are more likely to sustain the gains.

Capital-market conditions also matter. Periods of abundant venture funding can accelerate product development and reduce the price of commercial solutions, lowering the adoption threshold for providers. Periods of tighter capital can slow the pipeline of new tools and force consolidation among vendors. Because many of the highest-value use cases require integration with existing electronic health records and claims systems, the bargaining power of the large platform vendors (EHR suppliers, major claims processors) will influence both the pace and the distribution of gains. Vertical integration between AI developers and care-delivery organizations is one possible response; open-interface requirements or interoperability mandates are another.

6.5 Non-Financial Benefits and Total Mission Value

Throughout the savings analysis Sahni et al. emphasize that many AI use cases generate value beyond the financial line. Quality and safety applications that detect deterioration earlier can avert adverse events whose human cost far exceeds the avoided medical spending. Capacity optimization that reduces waiting times improves patient experience and can expand effective access without new construction. Tools that reduce the administrative burden on clinicians—documentation assistance, automated prior-authorization support, smarter scheduling—can mitigate burnout and improve retention in a labor market already under strain. These effects are part of what the authors call total mission value: the sum of financial and non-financial returns that a mission-driven healthcare organization should consider. From an economic standpoint the non-financial benefits raise two related issues. The first is measurement. Unlike cost savings, which can at least in principle be observed in accounting data, improvements in experience or reductions in burnout require surveys, validated scales, or other instruments that many organizations do not routinely field. The second is appropriation. A hospital that invests in a deterioration-prediction system may capture some of the financial gain through lower length of stay or avoided penalties, but the pure safety benefit accrues to patients and, under many 17payment systems, is only partially reflected in the hospital's revenue. Public and private payers that care about these outcomes can close the gap by adjusting quality bonuses, safety metrics, or shared-savings calculations to credit the non-financial gains. Doing so would raise the private return to investments whose social return is already high.

There is also a dynamic interaction with adoption barriers. Organizations that evaluate AI projects solely on short-term financial ROI will systematically under-invest in use cases whose primary payoff is clinical or experiential. Those that adopt a total-mission-value frame, and that have the governance processes to act on it, are more likely to clear the investment hurdle and to sustain the complementary workflow changes. In that sense the measurement and appropriation of non-financial benefits are themselves organizational capabilities that affect the pace of diffusion.

Finally, non-financial benefits can feed back into the political economy of adoption. Clinicians who experience AI as a tool that reduces drudgery and improves the quality of the care they can deliver are more likely to become advocates than resisters. Patients who notice shorter waits, clearer communication, or fewer preventable complications may support further investment. Building those constituencies is part of the change-management task that the six organizational factors identified by Sahni et al. are meant to address.

7. Policy Implications

Several policy levers can raise the probability that the productivity gains materialize and are broadly shared. First, payment models that reward total cost of care and measured quality give provider organizations a clearer stake in cost-reducing, quality-improving AI. The growth of accountable care organizations, bundled payments, and capitated primary-care models already moves in this direction; further expansion would strengthen the incentive alignment.

Second, interoperability standards and public investment in data infrastructure lower the fixed cost of assembling training sets and of deploying models across organizational boundaries. The unfinished work of the 21st Century Cures Act and related rules remains relevant.

Third, clearer liability rules—whether through safe harbors for clinicians who follow validated AI recommendations in good faith, or through product-liability frameworks that place responsibility on developers when models are defective—would reduce the option value of waiting.

Fourth, workforce policy should anticipate the shift in task content. Training programs for both physicians and allied professionals can place greater weight on the complementary skills that AI does not easily replicate—complex communication, ethical reasoning under uncertainty, coordination of multi-morbidity care—while ensuring that clinicians can critically evaluate algorithmic outputs.

18Fifth, monitoring and evaluation capacity inside public programs (Medicare, Medicaid, Veterans Affairs) can generate the kind of transparent performance data that private markets often fail to supply, both accelerating learning and disciplining low-value products.

None of these measures requires central planning of technology choices. They aim instead to correct the incentive, information, and coordination failures that currently slow diffusion of tools whose social return appears to exceed their private return.

A further policy consideration concerns the non-financial returns that Sahni et al. deliberately leave unquantified. Improvements in patient safety, reductions in clinician burnout, and expansions of access in underserved areas are real social benefits even if they do not appear as line-item savings in a hospital's operating statement. Public payers and regulators that care about these outcomes can incorporate them into coverage and payment decisions—for example by rewarding reductions in adverse events or by adjusting quality measures to credit AI-enabled outreach that keeps high-risk patients out of the hospital. Doing so would help close the gap between private and social returns and would encourage investment in use cases whose primary payoff is clinical rather than purely administrative.

Antitrust and competition policy also intersect with AI adoption. If a small number of platform firms come to control the dominant clinical algorithms or the data networks on which they depend, the resulting market power could both slow diffusion (through high licensing fees) and distort the direction of innovation (toward tools that protect rather than disrupt incumbent rents). Maintaining contestable markets for AI tools, while still allowing the scale needed for data-intensive model development, is a delicate but important objective. Open standards, data-portability requirements, and careful merger review in the health-IT sector are the natural instruments.

Finally, the international dimension deserves brief mention. Other high-income countries face similar pressures of aging populations and rising healthcare costs, and several have invested heavily in national health-data infrastructures that in principle facilitate AI development. Comparative evidence on adoption rates, realized savings, and effects on clinician roles will be valuable for refining the U.S. policy mix. At the same time, differences in payment systems, professional regulation, and privacy regimes mean that lessons will not transfer one-for-one. The U.S. combination of fragmented payment, strong professional autonomy, and a large private sector creates both distinctive obstacles and distinctive opportunities.

8. Conclusion

The two papers that anchor this synthesis paint a coherent picture. The technological opportunity is real and quantitatively large: several hundred billion dollars a year in potential net savings from tools that already exist, with additional non-financial gains in quality, access, and experience. Realizing that opportunity, however, is not automatic. It requires organizational capabilities that many healthcare 19institutions still lack, industry-level coordination on data and standards, and a political economy that does not treat every incremental automation as an existential threat to professional status. From a labor-economics perspective the key insight is heterogeneity. AI will not affect all medical work equally. Tasks that rest on pattern recognition over structured data are more exposed to substitution; tasks that rest on soft information, relationship continuity, and contextual judgment are more likely to be complemented. The resulting reallocation of tasks and rents will reshape specialty choice, the division of labor between physicians and mid-level providers, and the internal organization of hospitals and groups. Those distributional consequences will in turn feed back into the politics of adoption.

The historical record of utilization review and evidence-based medicine offers a cautionary parallel. Tools that improve decision quality while leaving clinicians a meaningful residual role have diffused more successfully than tools that appear to replace judgment outright. AI systems designed with that lesson in mind—transparent, locally adaptable, and positioned as decision support rather than decision replacement—stand a better chance of crossing the adoption chasm.

Finally, the productivity stagnation that has characterized healthcare for decades is not an immutable fact of nature. It is the product of incentives, institutions, and the limits of earlier technologies. AI relaxes some of those technological limits. Whether the sector responds with higher productivity and broader access, or with another round of cost growth and defensive professional politics, will depend on choices that are still open.

The central economic question is not whether AI can perform useful tasks in healthcare—it already can—but whether the institutional environment will allow those tasks to be scaled in ways that raise social surplus and distribute it in a politically sustainable fashion. The savings estimates of Sahni and colleagues provide a quantitative benchmark for what is at stake. The task-based analysis of Dranove and Garthwaite supplies a framework for thinking about who will win and lose along the way. Together they suggest that the next decade of AI in U.S. healthcare will be shaped less by the raw capabilities of the algorithms than by the slower variables of payment design, data governance, professional norms, and organizational capacity. Those are variables that policy and management can influence. The opportunity is large; the outcome is not predetermined.

Future empirical work should prioritize three areas. First, rigorous evaluations of live AI deployments that measure not only model performance but downstream clinical and financial outcomes, ideally with quasi-experimental or experimental designs. Second, detailed task inventories and time-use studies that allow more precise mapping of specialty exposure to automation. Third, analysis of the general-equilibrium feedbacks—volume responses, wage adjustments, entry, and patient sorting—that will determine how large the net social saving ultimately is. Until that evidence accumulates, the prudent stance is cautious optimism: the technological opportunity is real, the barriers are equally real, and the net effect will depend on choices that are still being made. 20A brief note on the broader research agenda is in order before closing. The two papers that motivate this synthesis are part of a larger NBER project on the economics of artificial intelligence in healthcare.

Other contributions in the same volume address complementary questions—data availability and privacy, regulatory design, the ethics of algorithmic decision-making, and the role of management practices in technology adoption. Reading the present analysis alongside those papers would give a fuller picture of the constraints and opportunities. The present paper has focused on the productivity and labor-market margins because those are where the quantitative estimates and the task-based framework speak most directly. The institutional and ethical margins are no less important for the ultimate social value of the technology.

In practical terms, health-system leaders and policymakers who wish to move from the potential documented here to realized gains might begin with three concrete steps. First, inventory existing AI pilots against the domain frameworks in Sahni et al., asking which high-value use cases remain unaddressed and which organizational prerequisites (talent, data, change management) are missing. Second, map the task content of the clinical services that account for the largest share of local spending or quality shortfalls, using the routine/non-routine distinction as a rough screen for where AI is most likely to complement or substitute. Third, align payment and internal incentive structures so that the units that bear the cost of implementation also capture a meaningful share of the resulting savings or quality gains. None of these steps is sufficient on its own; together they raise the probability that the next five years deliver a measurable fraction of the opportunity that the research has identified. The stakes are high enough to justify the effort. Healthcare already absorbs nearly one-fifth of U.S. GDP. Even a partial realization of the 5–10 percent savings range would free resources for other social priorities, ease pressure on public budgets, and potentially improve the experience of patients and clinicians alike. The technology is no longer the binding constraint. The binding constraints are organizational, institutional, and political. Addressing them is the work that remains.

References

1. Acemoglu, Daron, and Pascual Restrepo. 2020. "Robots and Jobs: Evidence from US Labor Markets." *Journal of Political Economy* 128(6): 2188–2244.
2. Agrawal, Ajay, Joshua Gans, Avi Goldfarb, and Catherine Tucker, eds. 2024. *The Economics of Artificial Intelligence: Health Care Challenges*. Chicago: University of Chicago Press.
3. Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review* 53(5): 941–973.
4. Autor, David H. 2014. "Polanyi's Paradox and the Shape of Employment Growth." NBER Working Paper 20485.
5. Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118(4): 1279–1333.
6. Bates, David W., et al. 2021. "The Potential of Artificial Intelligence to Improve Patient Safety: A Scoping Review." *npj Digital Medicine* 4: 54.
7. Berwick, Donald M., and Andrew D. Hackbarth. 2012. "Eliminating Waste in US Health Care." *JAMA* 307(14): 1513–1516.
8. Cam, Alice, Michael Chui, and Bryce Hall. 2019. "Global AI Survey: AI Proves Its Worth, but Few Scale Impact." McKinsey Global Institute.
9. Cutler, David, Jonathan Skinner, Ariel Dora Stern, and David Wennberg. 2019. "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Spending." *American Economic Journal: Economic Policy* 11(1): 192–221.
10. Donaldson, Molla S., et al., eds. 2000. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press.
11. Dranove, David. 1988. "Demand Inducement and the Physician/Patient Relationship." *Economic Inquiry* 26(2): 281–298.
12. Dranove, David. 1993. "The Economic Evolution of Medical Practice." Various working versions; see also related discussions of utilization review.
13. Dranove, David, and Craig Garthwaite. 2022. "Artificial Intelligence, the Evolution of the Healthcare Value Chain, and the Future of the Physician." NBER Working Paper 30607. (Also Chapter 1 in Agrawal et al. 2024.)
14. Fuchs, Victor R. 1974. *Who Shall Live? Health, Economics, and Social Choice*. New York: Basic Books.
15. Gottlieb, Joshua D., et al. 2020. "Who Values Human Capitalists' Human Capital? The Earnings and Labor Supply of U.S. Physicians." NBER Working Paper 27196.
16. Gray, Bradford H., and Marilyn J. Field, eds. 1989. *Controlling Costs and Changing Patient Care? The Role of Utilization Management*. Washington, DC: National Academy Press.
17. Mullainathan, Sendhil, and Ziad Obermeyer. 2021. "Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care." *Quarterly Journal of Economics* 137(2): 679–727.
18. Rajkomar, Alvin, Jeffrey Dean, and Isaac Kohane. 2019. "Machine Learning in Medicine." *New England Journal of Medicine* 380: 1347–1358.

19. Sahni, Nikhil R., George Stein, Rodney Zemmel, and David Cutler. 2024. “The Potential Impact of Artificial Intelligence on Health Care Spending.” In *The Economics of Artificial Intelligence: Health Care Challenges*, edited by Ajay Agrawal et al., 49–75. Chicago: University of Chicago Press. (Earlier version: NBER Working Paper 30857.)
20. Sahni, Nikhil R., et al. 2019. Related analyses of U.S. healthcare productivity and administrative costs. McKinsey and academic collaborations.
21. Sahni, Nikhil R., et al. 2021. Analyses of administrative spending shares and industry structure.
22. 22Wennberg, John, and Alan Gittelsohn. 1973. “Small Area Variations in Health Care Delivery.” *Science* 182(4117): 1102–1108.
23. Wickizer, Thomas M., and Daniel Lessler. 2002. “Utilization Management: Issues, Effects, and Future Prospects.” *Annual Review of Public Health* 23: 233–254. 23