PitchBook
a Morningstar company

Institutional Research Group

Brian Wright
Lead Research Analyst,
Healthcare
brian.wright@pitchbook.com

Ben Zercher
Senior Research Analyst,
Biotech & Pharma
ben.zercher@pitchbook.com

Ben Riccio
Research Analyst,
Industry & Technology Research
ben.riccio@pitchbook.com

Adi George
Associate Data Analyst

pbinstitutionalresearch@pitchbook.com

Published on April 23, 2026

Contents

EMERGING TECH RESEARCH

AI Will Deliver Care to Billions and Break the System That Built It

PitchBook is a Morningstar company providing the most comprehensive, most accurate, and hard-to-find data for professionals doing business in the private markets.

Key takeaways

AI is already pervasive in care delivery on both the consumer and physician sides of the equation. We envision a world where cognitive diagnosis functions are near free, yet the downstream effects are likely to drive higher medical cost trends in the near term before seeing improved medical outcomes and lower costs in the long term. The massive cost burden of higher near-term AI-induced utilization (five to seven years) will bring the system to a financing breaking point with the cumulative increase in medical costs reaching $5.9 trillion by the late 2030's. Reductions in the administrative component of healthcare provider reimbursement (AI-driven administrative efficiencies are already being enacted) will provide the bridge financing and allow time to develop, test, and implement next generation payment models with Health and Human Services' various ARPA-H program models as early test cases. We believe next-generation outcomes datasets will need to be created to maximize the potential benefit of AI in care delivery, and next-generation outcome-based payment models will be developed.

The near-term value of physician cognitive functions increases with AI-induced higher utilization, but scenarios exist where it is significantly arbitraged away long-term, as AI exceeds human abilities, and care deserts are the wedge to get in the door. Next generation outcome-based measurements are underappreciated areas of investment focus, whereas traditional value-based care model investments are living on borrowed time. As true outcome measures are developed and tracked, AI-native behavioral modification technologies will be significant in improving clinical outcomes, and companies such as Lirio will be instrumental in improving healthcare outcomes. Lastly, the degree of improvement that is derived from the transformation of care delivery will be capped in societies that do not modernize health insurance underwriting regulations to fully incentivize healthy behaviors among consumers. As a result, these societies will continue to be at significant cost disadvantages in global labor markets.

We believe fragmentation of the delivery system, combined with the irreducibly personal nature of healthcare, will produce multiple winning care delivery models—not the winner-take-all dynamic typical of Big Tech platforms.

We advance eight proprietary thesis points that thread throughout this report:

1. **We believe fragmentation of the delivery system, combined with the irreducibly personal nature of healthcare, will produce multiple winning care delivery models**—not the winner-take-all dynamic typical of Big Tech platforms. Provider organizations such as Doctronic and Amazon Health that effectively embed AI-native capabilities across administrative and clinical capabilities will solve the age-old practice-scaling problem and allow physicians to perform above the top of their licenses.

2. **On the economic trajectory, we project that utilization will go higher in the near term and lower in the long term**, while costs will increase more than utilization in the near term and decrease more than utilization in the long term. However, the dynamics of higher utilization in early years will likely result in payment reforms, as even under the most aggressive clinical AI displacement scenarios, cumulative NHE remain above baseline through 2040 with a peak at $5.9 trillion in the late 2030's because the induced demand overhang is too large. Additionally, potential long-term utilization improvements will likely be viewed as theoretical, given near-term budgetary pressures. Administrative reimbursement is the lever that can provide bridge financing for this higher utilization. We provide full multi-scenario NHE projections through 2040 under multiple scenarios in the Appendix section.

3. **Big Tech will be more successful in healthcare long term than it has been historically**, driven by vertical integration strategies exemplified by Amazon Health AI and Google's clinical partnerships.

4. **Last-mile execution is the key differentiator for success**—the companies, such as Sprinter Health, that solve the final connection between AI capability and patient outcome will capture disproportionate value.

5. **Personality matters:** Clinical AI that lacks a human-centered design ethos will fail regardless of its technical sophistication.

6. **Winners will understand that the US government cannot fund healthcare at historical levels** and will build operating and financial models that recognize this fiscal reality. Hospital systems that reduce capital intensity—shifting from billion-dollar campus investments to distributed, technology-enabled delivery networks—will survive and thrive.

7. **The Chinese and Indian regulatory frameworks provide an advantage in adoption speed.** US regulatory friction will result in slower near-term adoption rates. However, clinician supply constraints—with a projected shortage of up to 124,000 physicians by 2034 and more than 83 million Americans living in primary-care deserts—will ultimately force the removal of regulatory friction. Initial US adoption will focus on lower clinical complexity tasks such as prescription refills, and adoption rates will be highest in areas with the greatest supply shortages.

8. **The ultimate bottleneck to a healthier society is a broken healthcare financing system in the US,** where providers are not incentivized to provide preventive care and consumers' incentives for behavioral modification are limited by regulations.

Executive summary

The current state of AI clinical capabilities is that AI performs extremely well on complex reasoning tasks yet breaks down under uncertainty with missing information or changing context. This was documented extensively by Stanford's ARISE research and the broader Stanford HAI ecosystem. However, with each study highlighting errors or weaknesses in past clinical decision models, the pace of iterations and improvement vastly outpaces the abilities of studies to keep up with the latest model iterations, especially with foundational models writing their own code.

That boundary is moving fast. AI capabilities are advancing at an accelerating rate. In structured clinical tasks—diagnostic pattern recognition, treatment protocol matching, and drug interaction analysis—AI already supersedes human clinical capabilities. Within the near future, that superiority will extend to broader, less defined clinical parameters as foundation models improve their ability to reason under uncertainty. Super intelligence based on certain tests would indicate that we could already be there and by other definitional tests, AI will be there in the blink of an eye. As such, one must now contemplate a world state that lies beyond medical super intelligence.

This report examines five interconnected dimensions of AI in care delivery:

AI adoption in healthcare is pervasive and accelerating. 81% of US physicians now use AI in their practices, more than double the rate just three years ago. Consumer engagement with AI-generated health information is similarly widespread, with 79% of US adults likely to search online for health answers and a majority finding AI-generated responses at least somewhat reliable.

The care delivery technical stack is maturing. We map a five-layer stack—from foundation models and world models at the base, through data liquidity, workflow optimization, and care orchestration, to direct care delivery models and consumer health tools at the top—and profile the early players competing at each layer.

Data quality is the key to winning, and outcome measures must evolve. The US healthcare measurement infrastructure—built around Healthcare Effectiveness Data and Information Set (HEDIS), Centers for Medicare & Medicaid Services (CMS) quality programs, and retrospective population-level metrics—is structurally incompatible with the real-time, individualized interventions that AI generates. Solving this measurement gap is a prerequisite for scalable, reimbursable AI-enabled care.

The regulatory landscape is in flux. The FDA's January 2026 clinical decision-support (CDS) guidance widens the pathway for nondevice clinical decision support, but the most ambitious AI applications—autonomous diagnosis, agentic care management, robotic surgery—face a complex, multijurisdictional regulatory environment that will evolve over the next decade. Several states, including California, Texas, Illinois, and New York, are enacting or proposing sweeping AI transparency and consent laws that create material compliance obligations for healthtech companies. However, care deserts provide the wedge.

The vision for future care delivery is a structural transformation, not an incremental improvement. Care settings will continue migrating from inpatient to outpatient to home. Staffing models will evolve from physician-centric to team-based to AI-augmented to selectively autonomous. Robotic and microrobotic interventions will extend the frontier of what can be treated outside traditional facilities.

AI in healthcare delivery Is pervasive

Physician adoption

The American Medical Association's (AMA's) 2026 Physician Survey on Augmented Intelligence, released in March 2026, provides the most comprehensive snapshot of physician AI adoption to date. The survey of nearly 1,700 physicians across specialties, practice settings, and career stages found:

81%

of physicians now use AI in their practice

This is more than double the 38% rate reported when the AMA first surveyed doctors in 2023.[1]

2.3

The average number of AI use cases per physician

Up from 1.1 in 2023.[2]

75%+

of physicians believe AI improves their ability to care for patients, up from 65% in 2023.[3]

40%

of physicians report balanced attitudes of equal excitement and concern over AI use

Citing patient privacy and the integrity of the patient-physician relationship as top concerns.[4]

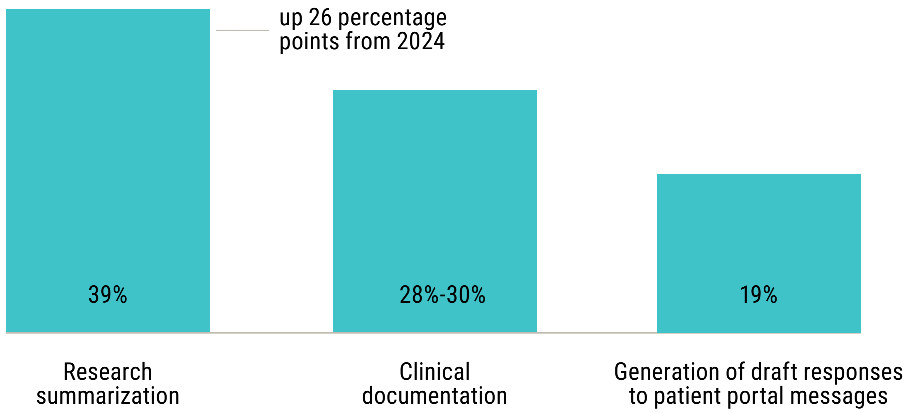
88%

express concern about potential AI-related skill loss

70%

are specifically worried about the impact on medical students and residents currently in training.[5]

Most common AI use cases amoung surveyed physicians[6]



70%

view AI as a tool to automate tasks that contribute to burnout.[7]

For investors, the AMA data confirms that physician adoption has crossed from early-adopter to mainstream. The relevant question is no longer whether physicians will use AI, but which AI tools will become embedded in clinical workflows deeply enough to generate switching costs and recurring revenue.

Consumer attitudes

On the consumer side, the Annenberg Public Policy Center's April 2025 health survey of more than 1,600 US adults reveals a population that is actively engaging with AI-generated health information—and broadly trusting it.[8]

79%

US adults who say they are likely to look online for the answer to a question about a health symptom or condition[9]

66%

of those who search online report having encountered AI-generated responses (such as Google's AI Overview orBing's Copilot Answer) at the top of their search results

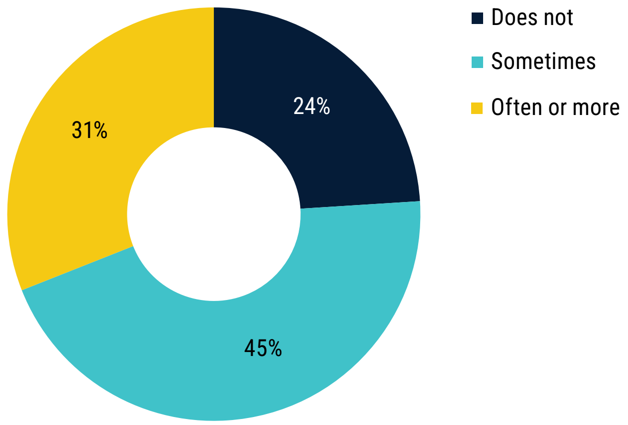
90%

of respondents still trust their physician as a source of health information

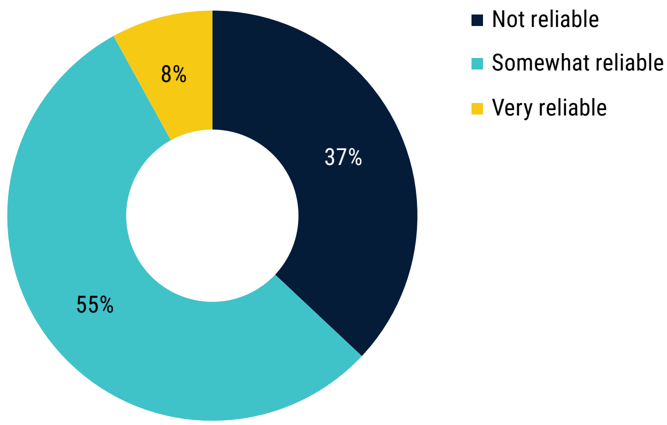
49%

are not comfortable with healthcare providers using AI tools rather than their experience alone when making decisions about their care.[11]

Do AI generated responses provide you with the answer you need?[10]



How reliable do you consider AI generated health information?[12]



This duality—consumers trust AI for information but prefer humans for decisions—has direct implications for care-delivery-model design. The winning models, in our view, will integrate AI seamlessly into the information and triage layers while preserving visible human clinical judgment for care decisions. This is precisely why we believe personality matters and the personal human touch will differentiate winning models.

for care-delivery-model design. The winning models, in our view, will integrate AI seamlessly into the information and triage layers while preserving visible human clinical judgment for care decisions. This is precisely why we believe personality matters and the personal human touch will differentiate winning models.

Benchmark results

AI clinical performance benchmarks are improving at a pace that consistently outstrips expectations, and AI now scores better than humans in most well-defined clinical functions.

HealthBench (OpenAI) represents the most comprehensive evaluation framework for assessing LLM performance on clinical tasks, spanning diagnostic reasoning, treatment planning, patient communication, and safety.[13] MedAgentBench (Stanford) extends evaluation beyond knowledge testing into agentic clinical task completion—assessing whether AI systems can not only reason correctly but execute clinical workflows (ordering labs, writing prescriptions, documenting encounters) within simulated electronic health record (EHR) environments.[14] Published research in Nature Medicine, the National Library of Medicine, and MIT economics has documented AI diagnostic performance that matches or exceeds physician accuracy across multiple clinical domains, with the gap widening as models improve.[15]

A 2026 study published in Nature Medicine evaluated ChatGPT's performance on structured triage recommendations, finding performance levels that suggest AI triage systems have not yet reached clinical-grade reliability, although given the pace of improvement, acceptable triage capabilities will likely be seen before more complex diagnostic tasks.[16] We would note that the structure of the study has been questioned by some experts.

Cost improvements in clinical tasks performed by AI are substantial and accelerating: As model inference costs decline and task-specific fine-tuning improves efficiency, the economic case for AI-assisted clinical workflows strengthens and contrasts with human labor, which becomes more expensive.

Reducing medical errors and establishing new standards of care: According to Dr. Ami Parekh, chief health officer at Included Health,[17] two of the most underappreciated aspects of AI on care delivery is the potential to reduce medical error rates and to define new standards of care. As administrative tasks are performed by agentic systems, clinicians and other care providers can focus their cognitive functions on the patient and clinical tasks. Included Health uses AI in its quality assessment program, which allows the company to review every clinical interaction with patients with an automated dashboard to categorize each interaction as green, yellow, or red; interactions receiving a red designation are flagged and receive human review. Beyond quality assessment, we see large language models' (LLMs') ability to process vast amounts of unstructured data and combine with traditional AI & ML capabilities on structured data has the potential to redefine standards of care and improve adoption rates.

Current AI-first care delivery models

AI agent use across the care delivery landscape

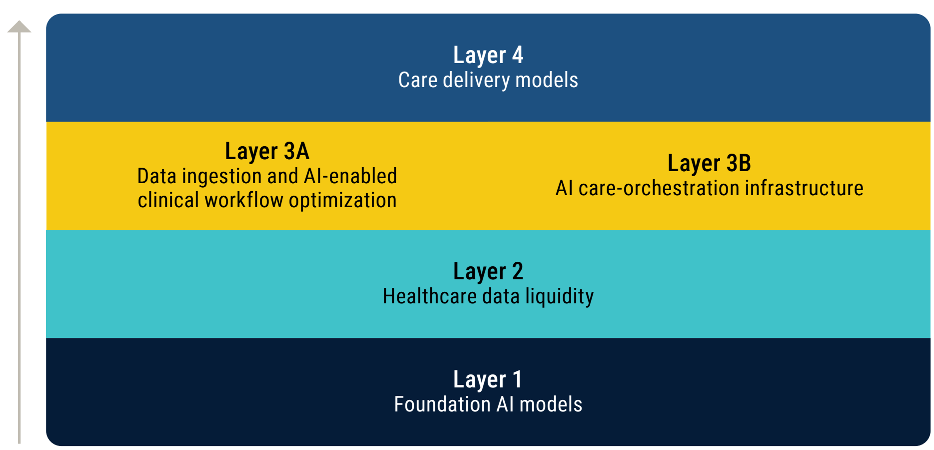
Front-office operations	Mid-office functions	Back-office functions
Scheduling	Clinical documentation	Billing
Patient intake	Prior authorization	Coding
Insurance verification	Care coordination	Claims management

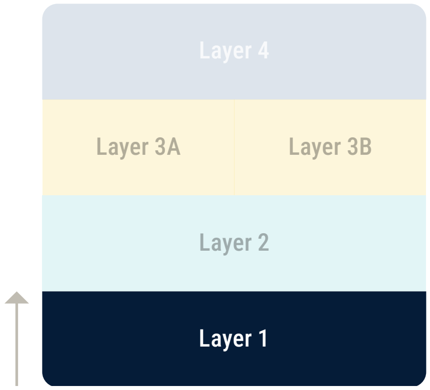
Initial integrated business models are developing across the care delivery landscape. AI agents are improving front-office operations (scheduling, patient intake, and insurance verification), mid-office functions (clinical documentation, prior authorization, and care coordination), and back-office functions (billing, coding, and claims management)—as we have discussed in prior reports on AI scribes and AI healthcare services agents.

Importantly, care-delivery-model innovation is not confined to the VC- and PE-backed universe. Private practices and hospital systems are actively ideating evolving care delivery models that leverage AI capabilities within their existing clinical operations. This breadth of innovation—from well-funded startups to community health centers—supports our thesis that fragmentation will produce multiple winning models rather than a single dominant platform.

AI clinical-care delivery stack

We conceptualize the AI care delivery ecosystem as a five-layer stack, with each layer building on the capabilities of the layers below it. This framework allows investors to evaluate where value accrues, where competitive moats exist, and where companies are positioned for durable advantage versus commoditization.





Layer 1—Foundation AI models

At the base of the stack sit the LLMs and AI infrastructure powering healthcare applications. These are the general-purpose reasoning engines upon which all higher layers depend. Below are recent healthcare announcements for each:

OpenAI introduced ChatGPT Health, which allows users to connect medical records and fitness data (via a back-end partnership with b. well). ChatGPT Health is designed to support medical care awareness and is not intended for diagnosis or treatment. It helps consumers “navigate everyday questions and understand patterns over time,” enabling patients to be “more informed and prepared for important medical conversations.”[18] In addition, OpenAI introduced HealthBench, a new benchmark designed to better measure the capabilities of AI systems for health, and GDPval, a task-based evaluation benchmark; both tools draw data from real-world scenarios.

Google released MedGemma 1.5 4B. Last year, Google published the MedGemma collection of open medical generative AI models, which are intended to be starting points for developers to evaluate and adopt medical use cases. This model update includes high-dimensional medical imaging, longitudinal X-ray imaging, anatomical localization X-ray imaging, and structured data extraction from medical lab reports. Google also recently released MedASR, a new open automated speech recognition model fine-tuned for medical dictation.

Anthropic announced Claude for Healthcare, a HIPAA-ready enterprise suite targeting healthcare administration. The core of the announcement was the launch of “Connectors”—native integrations that allow Claude to pull live data from the CMS Medicare Coverage Database, PubMed, the National Provider Identifier Registry, and International Classification of Diseases 10. On the life sciences front, Anthropic added deep-bench tools for clinical trials, integrating with platforms such as Medidata and ClinicalTrials.gov.

xAI represents a newer entrant whose healthcare-specific strategy is less defined but whose compute infrastructure and model capabilities are rapidly scaling, and the company has a consumer retail focus.

For investors, the foundation model layer presents a familiar Big Tech oligopoly dynamic: massive capital requirements, winner-take-most economics, and value that is increasingly captured through vertical integration into domain-specific applications rather than through horizontal model licensing alone.

World models

World models represent a fundamentally different AI paradigm from text-based LLMs. Where LLMs need only to read and process words, world models need to see and feel physical reality—understanding spatial relationships, material properties, and the physics of biological systems. This requires a massive convergence of next-generation compute power and a vast array of physical and biological sensors.

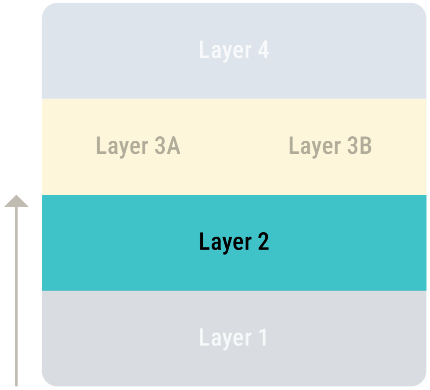
Compute Engine: NVIDIA Blackwell Ultra (and the forthcoming Rubin architecture): To process the physics of the real world, standard server infrastructure is insufficient: Gigawatt-scale AI factories are required. NVIDIA's Blackwell Ultra is designed for the multidimensional mathematics of physical AI and world models. Unlike the Hopper series, Blackwell Ultra is built to process high-fidelity multimodal data—video, 3D spatial mapping, and sensor data—simultaneously with drastically lower latency. This hardware works in tandem with the NVIDIA Omniverse platform to build digital twins, allowing AI to practice complex tasks, such as surgery, in a simulated environment or simulate chemical reactions before real-world deployment.

Training: Cameras and tactile sensors (MedOS): The recently launched MedOS, built by researchers at Stanford and Princeton, is an "AI-XR-cobot" system that trains embodied world models for clinical environments. MedOS uses extended reality (XR) smart glasses and clinical cameras to see the surgical world in 3D. Its foundational model was trained on "MedSuperVision," a dataset of over 85,000 hours of surgical video, allowing the AI to learn how tissues deform, bleed, and react to surgical tools. Vision alone is insufficient for surgery—the AI needs to understand resistance. MedOS's robotic arms use custom tactile sensors and force-torque monitors that feed continuous resistance data back into the world model, teaching it to distinguish healthy muscle from calcified tumor by physical "feel."

Microscopic training: Biological sensors (Bioptimus): Companies building world models at the cellular level, such as Bioptimus with its M-Optimus model, use advanced laboratory hardware—digital pathology scanners capturing massive H&E (tissue staining) images and multi-omics sensors (genomic sequencers, mass spectrometers)—to model how drugs physically alter cell structures over time. The world model fuses chemical sensor data with microscopic imaging data to create predictive biological simulations.

Training a world model is fundamentally about bridging the physical-to-digital divide. The sensors capture reality, and the Blackwell Ultra GPUs provide the brute force to turn that reality into a predictive simulation.

Advanced Machine Intelligence (AMI), co-founded by Yann LeCun, Michael Rabbat, and Alexandre Lebrun (Nabla co-founder), raised a $1 billion seed round in January 2026. AMI's goal is to build intelligent systems that understand the real world. According to its website, "AMI will advance AI research and develop applications where reliability, controllability, and safety really matter, especially for industrial process control, automation, wearable devices, robotics, healthcare, and beyond." The website further states that "Action-conditioned world models allow agentic systems to predict the consequences of their actions, and to plan action sequences to accomplish a task, subject to safety guardrails." Beyond Nabla's co-founder becoming the CEO of AMI while also serving as chief AI scientist and chairman of Nabla, Nabla announced in December 2025 an exclusive partnership with AMI to pioneer the next era of agentic healthcare AI. As a result of the partnership, Nabla will gain first access to AMI's world model technologies with a goal of becoming the first to bring FDA-certifiable agentic AI systems to healthcare.



Layer 2—Healthcare data liquidity

This layer comprises the interoperability platforms, healthcare data application programming interfaces (APIs), and clinical data infrastructure providers that allow AI systems to safely interact with fragmented healthcare data. Without this layer, even the most capable foundation model cannot access the clinical data it needs to reason effectively.

Data de-identification and linking: Essential for training models on real-world evidence (RWE) without legal risk.

- **Datavant** is the market-leading health data de-identification and linking platform, enabling organizations to connect de-identified patient data across disparate sources—payers, providers, labs, and pharmacies—to build longitudinal research and analytics data sets. Datavant's token-based linking technology has become an industry standard, with connections to thousands of data partners. For investors, Datavant's network effects create a compounding competitive moat: The more data partners connect through Datavant's tokens, the more valuable the network becomes for every participant.

- **HealthVerity** provides a complementary data de-identification and linking platform with particular strength in connecting real-world data for life sciences research and commercial analytics. HealthVerity's IPGE (Identity, Privacy, Governance, and Exchange) platform emphasizes privacy-preserving data linkage.

Real-time data APIs: Often described as the "plaid of healthcare," these companies allow AI agents to pull clinical data in real time during patient encounters.

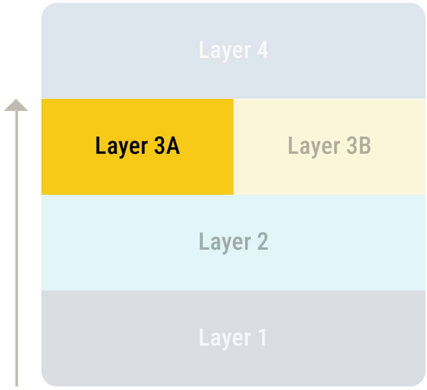
- **Zus Health** has built a shared health data infrastructure that aggregates clinical, claims, and social determinant data into a unified patient view accessible via API.

- **Particle Health** provides FHIR (Fast Healthcare Interoperability Resources)-based APIs that enable real-time access to clinical data from EHRs, health information exchanges, and payer systems.

Semantic normalization: AI cannot reason across "dirty" data. Inconsistent terminology, variable coding conventions, and unstructured clinical notes degrade model performance and increase hallucination risk.

- **Redox** provides a healthcare integration platform that normalizes data across disparate systems into standardized formats, enabling AI applications to connect with EHRs, labs, and other clinical systems through a single integration point.

- **b.well Connected Health** provides a consumer-facing health data aggregation and normalization platform that brings together clinical, claims, pharmacy, and wearable data into a unified patient health record.



Layer 3A—Data ingestion and AI-enabled clinical workflow optimization

This layer encompasses platforms focused primarily on automating specific workflows within health system operations. Ambient listening technologies, while addressed in prior reports, have evolved to become the data ingestion layer—whether offered by the EHR or independently.

Data ingestion layer:

- **Abridge** is the leading independent ambient clinical documentation platform, using AI to generate structured clinical notes from physician-patient conversations in real time. Abridge has secured partnerships with major health systems, including UPMC, Epic, and others, and its integration into Epic's EHR ecosystem gives it a significant distribution advantage. For investors, Abridge's position as the data ingestion layer creates a strategic option value: The ambient data it captures becomes the input layer for higher-level AI clinical reasoning.

- **Ambience (Ambience Healthcare**) provides an AI-powered ambient documentation platform with expanding capabilities in CDS and care coordination. Its "AutoScribe" product captures clinical encounters and generates notes, orders, and referral letters.

- **ModMed** (Modernizing Medicine) combines a specialty-specific EHR platform with AI-powered ambient documentation, creating a vertically integrated offering for specialty practices (dermatology, ophthalmology, orthopedics, and others).

- **Suki** provides an AI-powered voice assistant for clinical documentation, integrating with major EHR systems to reduce physician documentation burden. Suki's agent-based architecture allows it to perform tasks beyond documentation, including navigating EHR workflows and retrieving patient information.

- **Nabla** offers an AI copilot for clinicians that combines ambient documentation with CDS and serves both as a data ingestion tool and a Layer 1 adjacency through its foundation model partnership with AMI.

  Nabla's partnership with AMI marks a key step in its long-term effort to build a clinical AI platform capable of operating across the full complexity of care delivery. While today's large language models have unlocked significant gains in documentation and workflow support, Nabla believes the next phase of healthcare AI will require systems that can model clinical environments, reason over time, and safely support decision-making within real-world constraints. Through early access to AMI's world model technologies, Nabla is working toward capabilities such as auditable decision-making, simulation-based reasoning, and the handling of multimodal medical signals, which can support applications across areas such as workflow orchestration, clinical decision support, and eventually autonomous systems under clinician oversight. These advances are critical to enabling trusted, agentic systems in healthcare. The partnership with AMI also strengthens Nabla's positioning within the AI healthcare stack, aligning with its broader vision of a clinical intelligence layer that sits between fragmented systems and care teams, rather than a point solution.

Nabla's mission is to restore the human connection at the heart of healthcare by building the most advanced and reliable AI assistant for clinicians. By bringing the next generation of AI into healthcare, Nabla is helping build systems that understand not just language, but the realities of care itself. Ultimately, Nabla and AMI share a long-term ambition to advance toward AI systems that can operate safely and reliably in high-stakes clinical environments.

Clinical workflow optimization:

- **Notable** automates administrative and operational workflows across the patient journey—scheduling, intake, insurance verification, and referral management—using AI to eliminate manual steps and reduce staff burden.

- **Fabric** provides a digital front door and care-access infrastructure platform that uses AI to route patients to the right care setting, automate triage, and streamline the intake process.

- S**ullyAI** focuses on clinical documentation and workflow automation with emphasis on specialty-specific optimization.

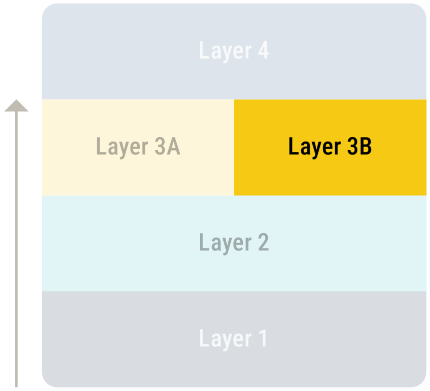
- **OpenEvidence** provides AI-enabled CDS, using evidence-based content to help clinicians make informed treatment decisions at the point of care. OpenEvidence partnered with Elsevier in the development of ClinicalKey AI.

- **Counsel Health** offers AI-enabled CDS with a focus on guiding clinicians through complex diagnostic and treatment pathways.

- **Hippocratic AI** is building AI-enabled patient interaction and care navigation agents—AI systems designed to communicate directly with patients for tasks such as appointment scheduling, medication reminders, pre-visit preparation, and post-discharge follow-up.

- **Latent Health** has developed a clinical reasoning model specifically for medication prior authorization and management—one of the highest-friction, highest-cost administrative workflows in US healthcare.

Layer 3B—AI care-orchestration infrastructure



Platforms designed to extend care teams and coordinate longitudinal care delivery across populations. These function more like clinical operating systems for care delivery, integrating into provider workflows and supporting care teams at scale.

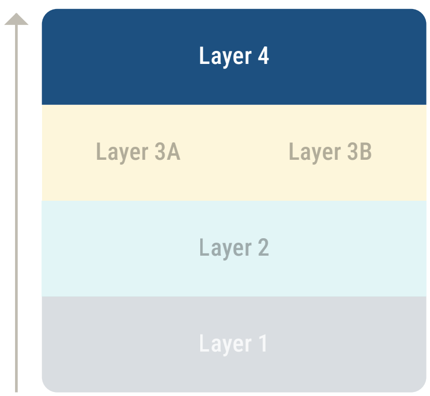
- **Lumeris** provides an AI-enabled primary-care infrastructure and care-orchestration platform that supports health systems and physician groups in enabling greater access to primary care, extracting more value from the service and improving experiences for patients and providers in the transition to value-based care. Lumeris combines technology, data analytics, and managed services to enable population-health management orchestration of ambulatory care at scale. We will be providing a company spotlight report on Lumeris, whose last-known valuation was $2.1 billion in 2024, in the near future, detailing the company's model.

- **Qualified Health** is a public benefit corporation that provides an end-to-end, enterprise-wide AI operating platform designed specifically for health systems. The platform centralizes the orchestration, deployment, and governance of generative AI, offering both pre-validated clinical/operational workflows and proprietary "builder tools" that allow hospital teams to safely create and scale custom AI agents.

- **RhythmX AI** offers AI-enabled care coordination and workflow orchestration, helping care teams manage patient panels, coordinate referrals, and track longitudinal outcomes.

- **Ellipsis Health** provides AI-enabled care management and coordination, with a particular focus on behavioral health screening and monitoring using voice biomarker analysis.

- **UpDoc** delivers AI-powered remote patient intervention, using predictive analytics and AI-driven outreach to engage patients between visits and prevent acute episodes.



Layer 4—care delivery models

Organizations delivering care directly through virtual or hybrid care models. There is a virtual graveyard of practice-based models that have tried to crack the code for advanced primary care, but they did not have the tools needed to expand the bandwidth of clinicians to allow them to scale more effectively. Generative AI changes that. We are seeing a new generation of care delivery applicants with significantly improved probability of success.

- **Doctronic**, a New York City-based AI-native healthcare startup, has positioned itself as the leading consumer-facing "AI doctor" platform in the US. Founded in 2023 by Adam Oskowitz (a practicing UCSF vascular surgeon) and Matt Pavelle (a repeat founder and former CTO of Moda Operandi), the company has achieved extraordinary traction in under two years—growing from zero to eight-figure annualized revenue, more than 300,000 unique weekly visitors, and more than 15 million medical conversations.[19] In December 2025, Doctronic became the first AI platform in US history to legally and autonomously renew prescriptions, a regulatory milestone achieved through Utah's AI Learning Lab sandbox. The company closed a $40 million Series B in February 2026 at an approximately $206 million post-money valuation, bringing total capital raised to $65 million.

- **K Health** is an AI-enabled digital primary-care provider that uses a symptom-checker AI trained on billions of clinical data points to provide personalized health assessments and connect patients with clinicians.

- **Akido Labs** is an AI-enabled digital primary and specialty care provider. CEO Prashant Samant has articulated the vision of physicians as "field generals" overseeing expanded care teams empowered by AI.

- **Circle Medical**, a virtual-first primary-care network, operates across multiple states, combining telemedicine with AI-powered clinical workflows.

- **Lotus Health**, and AI-enabled digital clinic, is focused on technology-driven care delivery.

- **Summa Health** is one of Ohio's largest integrated healthcare delivery systems, headquartered in Akron. Founded in 1989 through the merger of Akron City Hospital and St. Thomas Hospital, the system has grown into an approximately $2 billion annual revenue enterprise with 8,500+ employees. In October 2025, Summa Health became the first US hospital system ever acquired by a venture capital firm, when General Catalyst's spinoff—Health Assurance Transformation Company (HATCo)—completed a $515 million buyout/LBO, converting Summa from a nonprofit to a for-profit entity. Summa is ranked among Healthgrades America's 50 Best Hospitals, placing Summa in the top 2% of all hospitals nationally for clinical care and patient outcomes.

  This deal represents a paradigm shift in healthcare delivery investment, with implications for how VC/PE capital can be deployed to transform legacy health systems through technology, value-based care, and operational innovation. When acquired, General Catalyst stated that "The focus isn't on taking costs out... it's on putting innovation in (we intend to build modern, tech-enabled healthcare delivery platforms at scale, across all points of care)."[20]

- **Lyra Health** is a mental health provider using AI-enabled matching of mental health providers with patients, optimizing the therapeutic alliance through data-driven provider selection.

- **Amazon Health AI**—As detailed in Section 6, Amazon's Health AI assistant, integrated with One Medical and Amazon Pharmacy, represents the most vertically integrated care delivery model in the stack. Its March 2026 expansion to 200 million Prime members transforms it from a primary-care provider into a population-scale health AI platform.

- **Jivi.ai (India)** is an AI-enabled primary-care delivery platform operating in India. The personal AI Doctor app has 3.5 million installs.[21] At the AI Impact Summit 2026 in India, Vinod Khosla indicated that all 1.5 billion people in India can be treated by AI doctors 24 hours a day/seven days a week at minimal cost.[22]

- **China's AI Hospital—Agent Hospital,** developed by Tsinghua University's Institute for AI Industry Research (AIR) and first published on arXiv in May 2024, is a virtual simulation environment rather than a physical hospital that treats real patients. All patients, nurses, and doctors within the system are LLM-powered autonomous agents that interact in a digital twin of a hospital setting.[23] At its core is MedAgent-

Zero, a self-supervised training framework in which AI doctor agents learn by treating synthetic patient agents, building a library of successful cases and reflecting on errors without requiring manually labeled training data. The initial training covered eight respiratory diseases, and after treating 10,000 virtual patients, the doctor agents achieved success rates of 88%, 95.6%, and 77.6% in examining, diagnosing, and treating patients, respectively.[24] The widely cited 93.06% accuracy figure comes specifically from performance on a subset of the MedQA benchmark (USMLE-style multiple choice questions) covering respiratory diseases—notably high compared to Med-Gemini at 91.1% and GPT-4 at 90.2%, although the comparison may not be apples-to-apples since the MedAgent-Zero result was achieved on a domain-specific subset.[25] As of April 2025, an inauguration ceremony announced the system's integration with Beijing Tsinghua Changgung Hospital, initially piloting services in general practice, ophthalmology, radiology, and respiratory care, but the platform does not yet operate as a physical hospital providing care to humans and would require extensive ethical review, regulatory approval, and further testing before real-world clinical deployment.[26,27]

- **DeepSeek in Chinese hospitals.** Running parallel to the Agent Hospital research effort is the bottom-up deployment of DeepSeek-R1, an open-source LLM, across China's hospital system. A nationwide cross-sectional survey found that DeepSeek-R1 had been locally deployed in 261 hospitals across 93.5% of Chinese provinces by early March 2025, with tertiary hospitals accounting for 84% of deployments and adoption concentrated in wealthier eastern regions.[28] On-premises deployment keeps patient data within hospital firewalls, and functional applications range from CDS and hospital administration to research facilitation and patient management, with reported efficiency gains including a 40-fold increase in patient follow-up throughput.[29] At Ruijin Hospital in Shanghai, an AI pathology model built on the platform processes roughly 3,000 slides per day, and other hospitals use it for triage, medical documentation, and chronic disease management. However, a JAMA perspective led by Wong Tien Yin, founding head of Tsinghua Medicine, warned that adoption may be "too fast, too soon," citing the model's tendency to generate plausible but factually incorrect outputs that could create substantial clinical risk, alongside concerns about clinician over-reliance and cybersecurity gaps at hospitals that lack adequate infrastructure.[30]

- **Advanced Research Projects Agency for Health (ARPA-H) ADVOCATE.** The federal government's most significant attempt to drive AI in clinical-care delivery is ARPA-H's ADVOCATE (agentic AI-Enabled Cardiovascular Care Transformation) program, announced in January 2026. ADVOCATE aims to develop the first FDA-authorized agentic AI technology capable of providing 24/7 specialty care for patients with advanced cardiovascular disease—the leading cause of death in the US, responsible for more than 200,000 preventable deaths annually.[31] The program's stated goal is nothing less than creating a regulatory and reimbursement blueprint that future AI clinical tools can follow. ARPA-H estimates the program could generate more than $50 billion in annual savings through optimized treatment for heart disease patients.[32]

  ADVOCATE is structured across three technical areas: a patient-facing clinical AI agent, a supervisory AI agent that monitors the clinical agent's ongoing safety and effectiveness, and a scalable integration plan for deployment across diverse care environments—from major health systems to rural clinics. ARPA-H is working directly with the FDA to chart a regulatory path, and with the Office of the National Coordinator for Health IT (ONC) to ensure data access for innovators.

- **University of Texas—Austin Medical AI Hospital of the Future.** The University of Texas system is building what may be the most ambitious AI-native healthcare facility in the world. Anchored by a $25 million state legislative appropriation and $100 million from Tench and Simone Coxe (January 2026), one of the largest gifts in UT history, the UT Research, Engineering, and Application Laboratory for Healthcare Artificial Intelligence (UT REAL Health AI) initiative is constructing the intellectual and operational infrastructure for a new kind of hospital—one designed from the ground up around artificial intelligence, robotics, and ambient sensor networks rather than retrofitting legacy facilities with bolt-on technology. The physical manifestation of this vision is the $2.5 billion UT Austin Medical Center, a two-hospital campus. We will be providing a closer look at UTA's hospital of the future in upcoming reports.

  For healthcare investors, this is not a single company to back but an ecosystem to understand. The technologies, platforms, and operational models being developed here will define the standard for how hospitals are built and operated for the next generation. The companies that become embedded in this ecosystem—whether as sensor providers, AI platforms, robotics manufacturers, or infrastructure builders—will have a reference installation. In the Appendix of this report, we provide a detailed analysis of this project.

Real-world comparison—current models

In an attempt to make lemonade, we used an unfortunate soccer injury to test existing models' clinical diagnostic capabilities as of early March 2026. We provide a summary of our impressions of the capabilities of the following consumer-facing AI-native care delivering models as of early March 2026. We list the overall experience in order of preference for each of the platforms.

Overall experience in order of preference

Model	Summary of our impressions
Doctronic	We view Doctronic as providing the best overall experience because the follow-up questions most effectively recreate questions we would expect from a specialist in the relevant medical field, which increased our confidence in the diagnostic assessment. The patient used its AI Doctor Consultation function, which is free of charge. The patient provided a brief description of the injury and resulting functionality. The AI Doctor then asked a series of follow-up questions regarding the specific location of the pain and overall pain levels on a scale of one to 10 to better ascertain the extent of the injury and provided a quick and thorough comprehensive medical assessment include smart medical triage. Additionally, Doctronic provided the patient with a subjective, objective, assessment, and plan note from the encounter, all at no charge. Had the patient wished to speak to a licensed medical professional, he could have requested a real-time visit with one of its staff physicians for the very reasonable price of $39. Given personal experience with similar prior injuries, the patient concluded the injury was an ankle sprain, which was consistent with the AI diagnosis; the patient did not request to see a physician. One minor critique of functionality was the lack of ability to upload a picture of the injury, whereas the LLM platforms offer that capability.

Overall experience in order of preference (continued)

Model	Summary of our impressions
Amazon Health AI	The advantage of the now-free Amazon Health AI agent is that with one click you can give access to your medical records to enable better clinical context for diagnoses. The drawback is that if you want to utilize the platform, you must provide consent for access to your medical records. In addition, under its introductory offer, Prime members get up to five direct message treatments with a medical provider via Health AI, subject to restrictions that include 30+ common conditions. This is beyond the free Ask Health AI agent function that includes a symptoms checkers, which is what we used. Beyond the introductory offer, Amazon, through its OneMedical asset, offers a $29-per-treatment charge. We note that the symptom checker went straight into diagnosis and treatment recommendation without asking any follow-up information in terms of pain levels and location of pain. One feature the patient appreciated was the ability to upload photos.
xAI	xAI offered the best all-around experience in terms of mix of bedside manner, diagnosis accuracy and treatment recommendations. xAI provided more detailed rationale on its diagnosis along with providing expectations on what to expect in terms of swelling and bruising over the coming weeks. As with the others, it offered the standard RICE treatment recommendation and provided a list of symptoms to monitor that it recommended would necessitate seeing a doctor sooner rather than later.
OpenAI	The ChatGPT experience was the best of the LLMs in that it was most descriptive in terms of interaction experience or bedside manner. In terms of diagnosing ankle sprain grade as likely between a Grade 1 and Grade 2, the patient would probably more agree with a Grade 2 diagnosis provided by the other LLM models. ChatGPT offered the standard RICE treatment recommendation, provided a typical recovery timeline and offered to 1) estimate the patient return-to-play timeline, give a day-by-day rehab plan, or 3) help tell the patient if it is safe to start jogging next week. Given the severity, the patient knew that jogging next week would not be possible.
Gemini	Offered a more concise response and probably a more accurate Grade 2 level ankle sprain and its updated version of RICE 1) Compression & Support, 2) Relative Rest, 3) Elevation, and 4) Early Range of Motion. However, the response felt less personal and more sanitary. Gemini then offered to 1) draft a list of questions for a doctor's visit, 2) search for local physical therapy clinics, or 3) create a four-week ankle rehabilitation plan.

Real-world comparison conclusion

At the end of the day, the Doctronic, AmazonHealth, and LLM models provided perfectly reasonable diagnosis and care guidelines for a lower-severity injury and likely saved the patient anywhere from $300 to $3,000 in out-of-pocket medical expenses for going to an ER or urgent care facility that would most likely require an X-ray and examination charges.

Consumer AI health tools outside the regulated care delivery stack

Many of these companies position themselves as health coaches, health advocates, or informational tools, engaging in clinically adjacent conversations without directly delivering regulated care. Additionally, this layer includes member/customer engagement companies that nudge patients toward desired behavioral changes to enhance health outcomes.

- **Thrive AI Health**—A joint venture between Thrive Global and OpenAI, building a personalized AI health coach that integrates peer-reviewed science with individual user data to deliver hyper-personalized behavior-change recommendations.

- **Noom**—The leading AI-powered weight management and behavior-change platform, using cognitive behavioral therapy principles delivered through AI-guided programs.

- **Included Health**—Employer-sponsor focused virtual care and navigation platform using AI for front-line member interactions, including care management, care navigation, billing support, and front-office operations.

- **Lirio**—A behavior-change AI platform that uses behavioral science and predictive analytics to deliver personalized health engagement messages at scale, partnering with payers and health systems.

Outcomes data limitations

The structural incompatibility

Healthcare's clinical outcome measurement infrastructure was built for a different era—one of episodic patient encounters, annual reporting cycles, and process-proxy accountability. AI is now operating in a fundamentally different paradigm: continuous, individualized, prospective, and often autonomous. The mismatch between these two realities is not a gap to be patched through incremental policy updates. It is a structural incompatibility that undermines the accountability, reimbursement, and trust architecture upon which AI-enabled care must eventually stand.

The current measurement paradigm is not AI-ready. HEDIS, CMS quality programs, and existing RWE frameworks were designed for retrospective, population-level, process-proxy measurement. AI systems generate real-time, individualized, outcome-linked interventions that existing measures cannot capture, validate, or attribute.

The majority of HEDIS and CMS quality measures track process compliance: Was a mammogram ordered? Was a follow-up appointment scheduled within seven days of hospital discharge? Was an HbA1c test performed? When AI systems can autonomously execute, schedule, or document process steps, the process measure loses its informational content. A health plan that deploys an AI scheduling tool to close HEDIS gaps may achieve improved measure performance without any corresponding change in patient outcomes. Conversely, an AI system that prevents hospitalizations through early intervention may not generate the process events that HEDIS measures track and will appear measurement-neutral despite generating substantial clinical value.

This dynamic presents a direct challenge for investors. We believe that accurate reporting and tracking of clinical outcomes will be the key differentiator for AI clinical model success over the long term. Companies that can demonstrate measurable improvements in patient outcomes—not merely process compliance—will command premium valuations and sustainable competitive moats.

The National Committee for Quality Assurance's) (NCQA's) own trajectory reflects the system's awareness of this problem. HEDIS Measurement Year 2026 specifications continue the transition from traditional hybrid reporting to electronic clinical data systems (ECDS) and FHIR-based digital quality measurement, with a mandate for all measures to be reported through ECDS by 2030.[33] Seven new measures were added for medical year 2026, including risk-adjusted outcome measures for acute hospitalizations following outpatient surgeries—a shift, however incremental, toward outcome-linked accountability. Still, these reforms remain squarely within the retrospective, population-level paradigm. None address the real-time, individualized attribution challenges that AI-generated interventions create.

US vs. state-mandated models

The US is at a structural disadvantage relative to centrally governed systems in China and India that can mandate reporting requirements and achieve near-total compliance. China's state-directed healthcare AI deployment, coordinated through its National Health Commission, can enforce uniform data collection standards and outcomes tracking across its hospital network. India's Ayushman Bharat Digital Mission similarly provides a centralized digital health infrastructure through which AI outcomes reporting could be mandated top-down.

The US advantage, however, lies in its marketplace's ability to innovate and develop better outcomes metrics than top-down regulations typically produce. The question for the sector—and for investors evaluating it—is how to incentivize adoption and standardization while continuing to innovate. We believe continuous refinement through agentic software development may hold the key. Agentic AI systems capable of autonomously collecting, normalizing, and reporting outcomes data in real time could bypass much of the manual reporting infrastructure that currently constrains the US measurement apparatus. ARPA-H's CARE ET chatbot program is an early example of AI interacting with patients in government health programs in the US.[34]

Incentive structures

Driving adoption of improved outcomes measurement will require both positive and negative incentives.

On the positive side—the carrot—financial incentives aligned with outcomes reporting can accelerate adoption. Value-based care contracts that tie reimbursement directly to AI-verifiable outcomes (rather than process compliance) would create natural demand for better measurement infrastructure. An emerging opportunity exists in shifting the burden of outcomes reporting to patients themselves, with AI navigation assistance and financial incentives making the reporting process frictionless. Consumer-facing AI tools that prompt patients to log symptoms, report functional status, and confirm medication adherence could generate continuous outcomes data at marginal cost—provided the patient experience is well-designed and the financial incentives are meaningful.

On the negative side—the stick—regulatory penalties for failure to report, or for submitting data that cannot be independently validated, can impose accountability. CMS's existing star ratings system, which directly impacts Medicare Advantage plan revenue through quality bonus payments, already provides a structural template. Extending similar frameworks to explicitly capture AI-generated intervention outcomes would force plans and health systems to invest in the measurement infrastructure.

We believe the winners in AI-enabled care delivery will be those that solve the outcomes measurement problem as a core competency—not as a compliance afterthought. This conviction threads through our broader thesis: Last-mile execution matters, and nothing is more "last mile" than proving that an AI intervention actually improved a patient's health.

Beyond ambient scribing—optical and sensor information

The ambient clinical documentation market—Abridge, Ambience, Suki, and others discussed in the AI clinical-care delivery stack—have established the principle that passive data capture during clinical encounters can generate structured, actionable clinical information. But ambient scribing captures only the verbal dimension of the clinical encounter. The next frontier is optical, sensor-based, and biometric data that expands the clinical information universe far beyond the spoken word.

For investors, this section represents a map of the data modalities that will feed the next generation of AI clinical models. Each new data source creates both a commercial opportunity and a competitive moat for companies that can ingest, normalize, and reason across it.

Phone cameras

The most ubiquitous medical imaging device in the world is the smartphone camera already in the patient's pocket. Dermatology triage applications that analyze skin lesions through phone-captured images are among the most commercially advanced AI clinical tools. Beyond dermatology, phone cameras are being used for wound monitoring in post-surgical care, dental screening, and even retinal imaging using clip-on adapters that transform the phone into an ophthalmoscope. Oral health screening through intraoral photography, gait analysis via video capture, and mental health assessment through facial expression recognition are all active areas of AI development leveraging the phone camera as a clinical instrument. The investor thesis here is straightforward: Phone-based imaging turns every consumer into a potential data source and every primary-care encounter into a higher-acuity diagnostic opportunity without additional hardware cost. The phone camera also serves as the most natural bridge between consumer health engagement and clinical-grade data collection—the patient does not need to acquire a new device, learn a new interface, or visit a facility.

In-exam, surgical, and hospital room capture

Within the clinical environment, camera-based data capture is rapidly moving beyond documentation into real-time CDS. In operating rooms, systems such as MedOS—the AI-XR-cobot platform developed by researchers at Stanford and Princeton—utilize extended reality smart glasses and clinical cameras to build 3D spatial maps of the surgical field. As described in the AI clinical-care delivery stack (Layer 1: World Models), MedOS was trained on over 85,000 hours of surgical video, allowing it to learn how tissues deform, bleed, and react to surgical instruments.[35]

In exam rooms, cameras paired with ambient AI can capture nonverbal clinical signals—gait analysis, facial expressions indicative of pain, and physical examination findings—that supplement the verbal data captured by ambient scribes. The combination of ambient audio (documenting the conversation) and ambient video (documenting the physical exam) creates a multimodal clinical record that is orders of magnitude richer than either modality alone. Hospital rooms equipped with continuous monitoring cameras can detect patient falls, monitor mobility, track sleep patterns, and identify early signs of delirium or agitation—all without requiring manual nursing documentation.

For investors, the optical data layer inside clinical environments represents a significant expansion of the total addressable market for ambient AI companies. Companies currently focused on ambient scribing (audio-only) face a natural product extension into ambient visual capture, and those that make this transition successfully will deepen their integration into clinical workflows and increase switching costs.

In-home diagnostics and preventive care

Sprinter Health provides on-demand, in-home diagnostic services and preventive care, now reaching patients across nearly 25 states following its February 2026 expansion of integrated Care+ visits.[36] Founded in 2021 by Max Cohen and Cameron Behar (both formerly of Oculus and Google), the company has raised more than $125 million—including a $55 million Series B led by General Catalyst with participation from Andreessen Horowitz, Google Ventures, Accel, and the Regents of the University of California.

Sprinter Health's model is built on a full-stack medical practice combined with a technology-first logistics platform. The company's field workforce—called "Sprinters"—are W-2 phlebotomists hired from the communities they serve and cross-trained in medical assistant and community health worker skills.[37] Using logistics AI that leverages route simulations, Sprinters are matched to patients within their own communities and dispatched to patients' homes for hands-on preventive diagnostics. During each visit, licensed and credentialed nurse practitioners connect virtually with patients to validate conditions based on objective findings, address higher-level clinical needs, and ensure each diagnosis is linked to a personalized, actionable care plan. Dedicated care navigators then connect members back to primary-care providers, health plan care teams, and community resources to ensure follow-through.

Portable ultrasound imaging: Within Sprinter Health's model and others like it, low-cost, handheld ultrasound devices—such as those from Butterfly Network—can be carried in a community health worker's bag and produce images interpretable by AI. This effectively democratizes a diagnostic modality previously confined to hospitals and imaging centers, enabling point-of-care imaging during home visits for conditions ranging from deep vein thrombosis to early pregnancy assessment to abdominal pathology.

Robot assistance (home health): Robotic assistants in the home health setting represent the convergence of multiple technology streams: autonomous navigation, computer vision, natural language processing, dexterous manipulation, and AI clinical reasoning. Current and near-term home health robots can monitor patient activity and mobility patterns, dispense and verify medication administration, provide physical therapy guidance through real-time motion tracking and feedback, collect ambient health data (temperature, air quality, activity levels), detect falls and alert caregivers or emergency services, and facilitate telehealth visits by providing a mobile video and audio platform that follows the patient through their home.

While fully autonomous home health robots capable of hands-on clinical care remain in the early stages, the trajectory is clear. The convergence of improved robotic manipulation (driven by advances in soft robotics and tactile sensing), computer vision (enabling real-time assessment of patient functional status), and AI clinical reasoning (interpreting multimodal data to generate personalized care recommendations) is narrowing the timeline to clinical deployment. According to perspectives shared by Geisinger leadership, 3% to 5% of the population consumes approximately 30% of healthcare resources at certain times—and these high-acuity, home-bound patients represent the highest-value target for robotic home health assistance.[38] For investors, home health robotics is a long-horizon opportunity with significant capital requirements, but the economics of the target population—high per-patient spend, labor-intensive care models, growing supply shortages in-home health aides—create a compelling total addressable market.

Investor implications

The proliferation of these data modalities across all three categories—optical capture, patient apps and wearable sensors, and home robotics—creates a secular tailwind for companies positioned at the data liquidity layer (Layer 2 of the stack) and the clinical workflow optimization layer (Layer 3A). Companies that can ingest, normalize, and clinically reason across multimodal data streams—not just text and voice, but images, waveforms, genomic sequences, and continuous biometric signals—will have structurally superior AI models and, consequently, better clinical outcomes. This is the data quality thesis: The model is only as good as the data feeding it, and the companies that solve multimodal data integration will build durable moats.

Payment models (reimbursement-based medicine)

Two competing philosophies within technology on care delivery reimbursement

The future of AI-enabled care delivery reimbursement sits at the intersection of two fundamentally different philosophies regarding how technology-driven medical care should be paid for—philosophies that will shape which business models succeed and which fail.

The first is that technology reduces cost, improves quality, and scales access. If AI can deliver a diagnostic recommendation that matches or exceeds a physician's accuracy, the reimbursement should reflect the value delivered to the patient, not the cost of the human labor it replaced. This philosophy favors capitated, outcome-based payment models that reward efficiency and quality regardless of the input modality—human or algorithmic.

The second is that as technology-driven care delivery input costs are substantially lower than traditional care delivery, the unit cost price advantage is passed on to the consumer, and the technology takes substantial market share. We believe long-term fiscal realities will require scaled cost improvement to be passed along to the consumer.

The medical industrial complex philosophy is built on the existing reimbursement infrastructure: fee-for-service billing codes, physician work relative value units (wRVUs), facility fees, and the elaborate coding architecture maintained by CMS, the AMA, and commercial payers. In this paradigm, reimbursement follows the provider, not the output. AI-generated clinical work that does not map to an existing CPT code, or that is not performed "incident to" a billable provider, may not be reimbursable at all—regardless of its clinical value.

The tension between these philosophies will define the business model viability of every company in the AI care delivery stack. Investors must underwrite to the reimbursement environment that actually exists while modeling toward the one that will exist in five to 10 years.

CMMI ACCESS program economics provide glimpse into long-term reimbursement mechanisms.

The CMMI ACCESS model: Overview

ACCESS—Advancing chronic care with effective, scalable solutions—is a new CMS Innovation Center model designed to expand options for people with Original Medicare and clinicians to manage chronic conditions with technology-supported care. The voluntary, 10-year program launches in July 2026 and is organized around four clinical tracks: early cardio-kidney-metabolic conditions (hypertension, dyslipidemia, obesity, prediabetes), cardio-kidney-metabolic conditions (diabetes, chronic kidney disease, cardiovascular disease), chronic musculoskeletal pain, and behavioral health (depression and anxiety).

Participating organizations receive recurring payments for managing patients' qualifying conditions, with full payment tied to achieving measurable health outcomes—such as improvements in blood pressure, HbA1c, lipids, weight, or validated patient-reported outcome measures for pain, mood, and function. This is what CMS calls "outcome-aligned payments," a deliberate departure from the traditional fee-for-service model that rewards volume of services rather than results.

Care can be delivered in person, virtually, or asynchronously, and participating organizations may leverage FDA-authorized devices, wearables, coaching apps, telehealth software, and other digital tools. ACCESS is designed to complement traditional care—primary-care physicians can refer patients to ACCESS organizations, receive regular electronic updates on patient progress, and bill a new co-management payment for reviewing those updates and coordinating care.[39]

The economics and the AI policy signal

The economics of ACCESS are revealing—and arguably intentional. The musculoskeletal track, which addresses one of the largest categories of healthcare spending in the country, will pay participants just $180 per beneficiary per year after the initial 12-month period for follow-on care. The model also includes a 50% payment holdback tied to meeting outcome targets, meaning organizations only earn their full reimbursement if patients achieve measurable clinical improvement. These payment levels are, by design, too low for traditional care delivery models staffed primarily by human clinicians operating in conventional workflows. CMS is setting reimbursement that only makes sense if most of the work is done by AI-driven companies capable of delivering care at significantly lower cost than traditional practice models. It looks to be a test case to see if AI-enabled care delivery can achieve measurable outcomes at dramatically reduced costs. While we believe this is the policy intent and long-term reimbursement paradigm, the healthcare industry has thrived off taking well-intended models and milking all the possible reimbursement avenues out of them, intended and unintended, and we see avenues to do just that under this structure.

Utilization impact of ubiquitous LLM access

The widespread availability of LLMs for health information will have complex and counterintuitive effects on healthcare utilization. Consider the patient who experiences a minor symptom—say, intermittent elbow pain after exercise. In the pre-LLM world, that patient might have ignored it entirely. Now consider the same patient querying an LLM. The AI might respond with a nuanced assessment and, in some cases, identify concerning patterns the patient would have dismissed, such as, "However, given your age, family history of cardiovascular disease, and the fact that you've mentioned occasional left arm numbness in a prior conversation, I'd recommend getting this evaluated promptly." The AI, operating with a lower tolerance for headline risk than the average consumer, will recommend clinical evaluation in scenarios where the patient would have otherwise self-triaged to inaction.

The net effect, in our view, follows a predictable trajectory:

Five-year horizon: Higher costs—The combination of increased diagnostic sensitivity (AI catches more), increased consumer health engagement (more people asking questions), and the persistent fee-for-service reimbursement architecture will drive utilization and costs upward. The AI functions as a highly effective demand generator within the existing payment structure.

Ten-year horizon: Lower costs—As AI-enabled preventive interventions mature, as chronic disease management tools demonstrate reductions in acute episodes and hospitalizations, and as payment models shift toward capitation and next-generation outcome-guaranteed payment models that reward prevention over treatment, costs begin to decline. The AI shifts from demand generator to demand moderator.

Long-term horizon: Lower costs—"It's the deficit, stupid"—Over the long term, the fiscal imperative will force the issue. The US federal deficit, driven significantly by Medicare and Medicaid spending, is on an unsustainable trajectory. Winners in AI-enabled care delivery will understand that the US government cannot fund healthcare at historical levels and will build operating and financial models that reflect this fiscal reality. This is not a policy preference—it is a mathematical constraint. We provide a thorough analysis of utilization and cost projections in the Appendix, section 3.

Deficit, unemployment, and physician shortages

The deficit trajectory is the ultimate forcing function. Federal healthcare expenditures as a share of GDP continue to grow, and the political appetite for additional taxation is limited.

Physician shortages compound the problem. The Association of American Medical Colleges (AAMC) projects a shortfall of between 37,800 and 124,000 physicians by 2034, with primary care bearing a disproportionate share.[40] Over 83 million Americans live in primary-care health professional shortage areas.[41] In medical deserts, AI-enabled care delivery is the only viable path to adequate access. We expect adoption rates to be highest in these underserved areas.

Personalized medications represent an additional cost vector in the near term but a cost reducer in the long term (fewer adverse drug events, fewer failed treatment attempts, better adherence).

Natural governors in healthcare

One of the underappreciated advantages of the healthcare sector is the presence of natural governors. As Robert Wachter eloquently states in his book, A Giant Leap, "While concerns about AI being deployed too rapidly are justified when it comes to society at large, healthcare's natural guardrails—our professional risk aversion, powerful incumbents, spring-loaded malpractice system, byzantine payment structures, and stringent privacy rules—produce ample amounts of brake-tapping. These constraints have created a healthy equilibrium—one in which we are gaining comfort with AI while implementing only reasonably well-vetted tools built by companies we trust."[42]

Regulatory framework

Software as a medical device (SaMD) vs. CDS

The regulatory architecture governing AI in clinical care pivots on a single, increasingly consequential distinction: whether an AI system is classified as SaMD subject to FDA premarket authorization, or as CDS software exempt from device regulation under Section 520(o)(1)(E) of the Federal Food, Drug, and Cosmetic Act.

On January 6, 2026, the FDA issued its updated final guidance on CDS software, superseding the September 2022 version.[43] FDA Commissioner Marty Makary framed the update as intended to reduce unnecessary regulation and promote innovation.

A CDS software function is excluded from device regulation only if it meets all four statutory criteria:

1. It does not acquire, process, or analyze medical images, IVD signals, or patterns from signal acquisition systems
2. It displays, analyzes, or prints medical information about a patient or other medical information
3. It provides recommendations to support, not replace, clinical judgment
4. It provides sufficient information for the clinician to independently review the basis for the recommendation.

The most substantive change in the 2026 guidance relates to Criterion 3. The 2022 version had effectively required CDS to present multiple options to avoid being classified as providing a "directive." The 2026 version introduces an enforcement discretion policy: If only one option is clinically appropriate, and the software otherwise meets all four criteria, the FDA will not enforce device requirements. This change materially expands the scope of AI-enabled clinical tools that can reach market without 510(k), De Novo, or PMA clearance.

The 2026 CDS Guidance also places greater emphasis on transparency for AI-driven tools, requiring that software or labeling provide plain-language descriptions of inputs, underlying logic, and how recommendations are generated. Software intended for time-critical decisions (stroke detection, sepsis alerts, patient deterioration monitoring) remains firmly within device regulation.

Importantly, as multiple legal analyses have noted, the guidance is silent on AI specifically. Commissioner Makary has publicly stated that the FDA is developing a new, forward-looking regulatory framework for AI, but that framework has not yet been published.[44]

The theoretical "software as a medical practitioner" (SaaMP) model

We outline a theoretical SaaMP licensing model requiring the following: knowledge and competency certification (including mathematical demonstration of calibration—proving the AI knows when it is uncertain); simulation-based clinical testing in high-stakes vignettes with ambiguous data, ethical dilemmas, and missing information; and operational accountability and auditing through continuous monitoring, mandatory incident reporting, and recurring audits. ARPA-H's PRECISE-AI program represents early federal investment in this accountability infrastructure.

The jurisdictional collision: FDA vs. state medical boards

The FDA is structurally equipped to evaluate a product at a point in time but ill-equipped for continuous practice oversight. State medical boards govern professional conduct but lack the technical expertise to audit neural networks. Resolution will require congressional action and coordinated overhauls of state medical practice acts. Until then, the management services organization (MSO)/friendly-PC structure and strict human oversight remain the only legally viable paths to market.

Liability: The unsolved "malpractice" problem

Currently, liability falls on the supervising clinician (automation bias) or the software developer (product liability). A fully licensed AI model would likely require specialized "algorithmic malpractice" insurance pools—a commercial opportunity for specialty insurers.

Joint commission and CHAI

The Joint Commission is developing AI deployment standards for accredited facilities. The Coalition for Health AI (CHAI) is emerging as an industry-led governance framework with guardrails, continuous monitoring, patient advocacy, and government cost advocacy.

The transition from AI as a "device" to AI as a "licensed practitioner" will require acts of Congress and coordinated overhauls of state medical practice acts. Until then, the MSO/friendly-PC structure and strict human oversight remain the only legally viable paths to market for these assets.

PCP deserts and the case for AI

Over 92 million Americans live in designated primary care health professional shortage areas.[45] As one Geisinger Health System executive framed the challenge: AI does not need to be perfect to improve the standard of care—it needs to be better than the status quo.[46]

Human-in-the-loop vs. autonomous

Healthcare is following a similar trajectory to aviation's progressive shift from manual to automated operations. According to Geisinger leadership, certain functions will always require human involvement, but others can be delegated to AI—not as "set it and forget it" but as continuously measured, checked, and monitored systems.

We project the US will retain human-in-the-loop requirements for five years, decanting lower-level functions (prescription refills, scheduling, routine lab interpretation) to autonomous operation.[47] Within 10 years, autonomous clinical decisions will obtain regulatory approval for well-defined, lower-risk scenarios.

Physician burden relief potential from AI adoption by specialty

Rather than addressing AI disruption risk by medical specialty, Gabe Wilson focused on how AI implementation could not only “give physicians hours back,” but also “transforms the quality of the hours they work.”

The analysis by Wilson, utilizing GPT 5.4-PRO, Gemini Deep Think, and Grok Heavy, projects the share of physician workflow AI meaningfully touches and estimates potential hours eliminated by specialty. Longer term, this could also be viewed as a framework for scale ability by specialty. A review of Wison’s methodology can be found here.

Functional lens: Radiology, internal medicine (general), dermatology, family medicine (general practitioner), and endocrinology have the highest level of AI assistance readiness and thus would generate the highest benefit from AI adoption while also facing higher long-term disruption risk. Anesthesiology, orthopedic surgery, and emergency medicine have the lowest level of AI assistance readiness and lowest long-term disruption risk.

Exhibit 1. AI assistable workflow by medical specialty

Specialty	AI readiness	AI-Assitable Workflow	Mean	Spread	Confidence	Key Bottlneck
Radiology (diagnostic)	High		77%	+/- 2.0	High	Medicolegal sign-off
Internal medicine (general)	High		64%	+/- 1.0	High	Multimorbidity complexity
Dermatology	High		63%	+/- 1.9	High	Hands-on exam / procedures
Family mMedicine / GP	High		62%	+/- 3.0	High	Broad undifferentiated scope
Endocrinology	High		62%	+/- 3.0	High	Complex titration / exceptions
Cardiology	Moderate		59%	+/- 6.4	Moderate	Unstable disease management
Psychiatry / behavioral health	Moderate		58%	+/- 1.5	High	Therapeutic relationship
Gastroenterology	Moderate		57%	+/- 4.0	Moderate	Endoscopy volume
Pediatrics	Moderate		56%	+/- 4.0	Moderate	Child exam / parent dynamic
Pulmonology	Moderate		54%	+/- 1.0	High	Diagnostic uncertainty
OB-GYN	Moderate		54%	+/- 5.0	Moderate	Pelvic exam / procedures
Neurology	Moderate		53%	+/- 3.0	High	Neurologic exam / localization
Urology	Moderate		51%	+/- 7.3	Low	Procedure dependence
Oncology (medical)	Moderate		50%	+/- 7.0	Moderate	Treatment selection risk
Ophthalmology	Low		49%	+/- 4.1	Moderate	Microsurgery / slit-lamp exam
General surgery	Low		47%	+/- 3.8	Moderate	Operative work
ENT / otolaryngology	Low		46%	+/- 5.0	Moderate	Specialized exam / scoping
Emergency medicine	Low		46%	+/- 2.0	High	Acute undifferentiated care
Orthopedic surgery	Low		46%	+/- 9.9	Low	MSK exam / surgery
Anethesiology	Low		45%	+/- 4.4	Moderate	Real-time airway control

Source: Gabe Wilson, GPT 5.4-Pro, Gemini Deep Think, Grok Heavy • Geography: Global • As of March 31, 2026

Exhibit 2. AI capability by functional role by medical specialty

Specialty	Mean	History	Diagnosis	Testing	Treatment	Procedure	Follow-up	Communication
Radiology (diagnostic)	77%	4.5	4.2	4.5	3.8	1.5	4.0	3.8
Internal medicine (general)	64%	4.2	3.4	3.6	3.5	1.6	3.9	3.8
Dermatology	63%	4.0	3.5	3.2	3.3	2.0	3.8	3.8
Family medicine / GP	62%	4.2	3.3	3.3	3.4	1.5	3.8	4.0
Endocrinology	62%	4.2	3.4	3.8	3.5	1.3	4.0	3.8
Cardiology	59%	4.1	3.3	4.0	3.2	1.5	3.7	3.6
Psychiatry / behavioral health	58%	4.3	3.2	2.8	3.0	1.2	3.5	3.8
Gastroenterology	57%	4.0	3.2	3.5	3.0	2.0	3.4	3.6
Pediatrics	56%	4.1	3.2	3.2	3.2	1.8	3.5	3.6
Pulmonology	54%	4.0	3.0	3.5	3.0	1.5	3.5	3.5
OB-GYN	54%	4.2	3.3	3.6	3.1	1.5	3.5	3.8
Neurology	53%	4.0	3.0	3.5	3.0	1.4	3.4	3.4
Urology	51%	4.0	3.1	3.4	3.0	1.6	3.3	3.5
Oncology (medical)	50%	4.1	3.2	4.0	2.8	1.4	3.9	3.7
Ophthalmology	49%	3.8	3.2	3.5	3.0	1.8	3.2	3.4
General surgery	47%	4.0	3.0	3.2	2.8	1.2	3.3	3.4
ENT / otolaryngology	46%	3.8	2.8	3.0	2.8	1.5	3.0	3.2
Emergency medicine	46%	4.0	2.9	3.5	2.7	1.6	3.1	3.4
Orthopedic surgery	46%	4.0	3.0	3.5	3.0	1.2	3.2	3.4
Anethesiology	45%	4.2	3.0	3.5	3.0	1.1	2.8	3.2
Dimension Average		4.1	3.2	3.5	3.1	1.5	3.5	3.6

THE PROCEDURAL WALL
1.5 Avg. Dimension E – the hard floor for AI in medicine

THE COGNITIVE CEILING
4.1 Avg. Dimension A – AI's strongest clinical domain

THE GAP
2.6 Cognitice assistant. Not physical replacement

Source: Gabe Wilson, GPT 5.4-Pro, Gemini Deep Think, Grok Heavy • Geography: Global • As of March 31, 2026

Care delivery visions for the future

Care-setting shifts

The secular migration of care from higher-acuity to lower-acuity settings—inpatient to outpatient to ASC to home—accelerates with AI. According to Geisinger leadership, 3% to 5% of the population consumes approximately 30% of healthcare resources at any given time. Remote monitoring with digital command centers represents the emerging operating model.[48]

Geisinger's strategic direction illustrates this trajectory: a slimmed presence through micro-locations rather than the colonial model of billion-dollar campus investments. Some small community hospitals, smaller than what were traditionally built, can meet the majority of a community's needs. This has profound implications for our thesis that hospital systems that reduce capital intensity will survive and thrive.

Staffing models

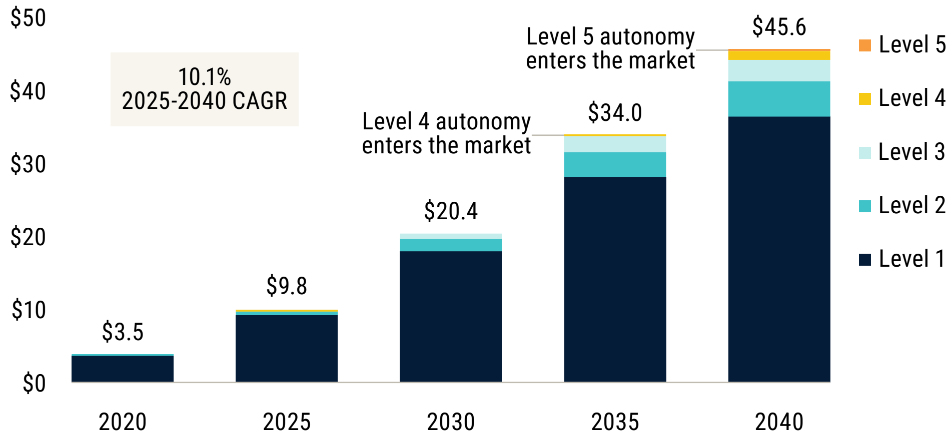
2036+ Beyond 10 years	AI handles routine encounters autonomously. Humans focus on complex cases, procedures, and the therapeutic relationship. Micro-robots begin transitioning from laboratory to clinical deployment.
2031-2036 Ten-year horizon	Autonomous AI systems handle well-defined clinical tasks. A new professional category—the AI supervisor—emerges. Fewer physicians per covered population, more technologists and AI system managers.
2026-2031 Five-year horizon	The physician remains the central clinical decision-maker, but AI dramatically expands their bandwidth. As Prashant Samant, CEO of Akido Labs, has described: "AI will amplify physicians' reach as they transition to being 'field generals' who oversee teams of medical assistants and healthcare technicians to provide their medical expertise to a much broader swath of patients."[49] Samant further notes that AI enables clinicians to focus on complex decision-making and relationship care, expanding panel size and increasing access. Marc Triola, director of NYU's Institute for Innovations in Medical Education, captures the pedagogical implication: the role of medical education must shift—"we should be teaching them how to conduct."[50]

Less than 20% of medical evidence currently makes it to patient care, with a 20-year translation gap. AI can collapse this gap dramatically—not by generating new evidence, but by ensuring existing evidence is applied consistently at the point of care. This is the distinction between innovation and adoption.

Robotic and micro-robot regulatory timeline

Autonomous robotic assistance for lower-risk procedures: five- to 10-year regulatory horizon. ARPA-H's AIR program targets autonomous thrombectomy, where more than half of Americans live 60+ minutes from a capable hospital and every 10-minute delay adds approximately $10,000 in costs.[68] Fully autonomous surgery: 20+ year horizon. Micro-robots: beyond 10 years for initial applications.

Exhibit 3. Robotic surgery market share history and projections



Source: PitchBook • Geography: Global • As of March 31, 2026

Fully autonomous robotic surgery will take at least 20 years to obtain regulatory approval in the US. ARPA-H's AIR program, targeting autonomous thrombectomy, represents the leading edge of this trajectory.[67]

Acknowledgements

One important disclaimer is that our contributors provided their insights and perspectives, but the conclusions of this report do not represent the opinions of these contributors. We did not discuss our conclusions with them as our discussions with them were during the fact finding phase of this report. This report benefited from the perspectives and insights of multiple contributors, and we would specifically like to thank the following for taking the time to speak with us over the prior months:

- **Byron Crowe,** chief medical officer, Doctronic
- **David G. Carmouche,** chief medical and commercial officer, Lumeris
- **Haider Warraich**, program manager | Scalable Solutions Office, Advanced Research Projects Agency for Health
- **Jensen Claus**, chief information and technology officer, University of Texas Medical Centers & Dell Medical School
- **Dr. Ami Parekh**, chief health officer, Included Health
- **Gabe Wilson**, Woodland Heights Medical Center
- **Ben Green**, chief clinical officer, AliveCor
- **Marten den Haring**, chief executive officer, Lirio

In addition, we are especially appreciative to Robert Wachter, professor and chair, Department of Medicine UCSF, whose book, A Giant Leap, was an instrumental element to our research. This book provided the most comprehensive published analysis of AI's impact on healthcare delivery. His discussions of the Joint Commission, CHAI, and institutional dynamics informed multiple sections.

We are also grateful to leadership at Geisinger Health System for sharing perspectives on care delivery transformation, the role of human oversight in AI-enabled clinical

workflows, the strategic shift toward micro-locations and distributed care delivery, the economics of high-acuity home-based populations, and the critical distinction between innovation and adoption in healthcare on the Healthcare on Air by Verizon series.

Appendix

The AI Cognitive Premium: An actuarial analysis of AI's impact on US healthcare expenditures, 2024–2040

Prepared by PitchBook March 2026 | Five-scenario model with dynamic cognitive premium and administrative automation variants

1. Executive summary

US healthcare spending reached $5.3 trillion in 2024. This analysis introduces the physician cognitive premium—the $600 billion to $950 billion annually paid for physician knowledge, judgment, and decision-making, exclusive of surgical procedures, diagnostic testing, imaging, rehabilitation, drugs, and facility costs—and projects how AI transforms this premium across five scenarios through 2040.

Our central finding: AI will increase costs for the US healthcare system from 2026 through approximately 2031 to 2033, as near-zero-cost AI diagnosis surfaces previously undetected conditions and drives downstream utilization on existing high-cost infrastructure. The cognitive premium itself is endogenous—rising from 15% to 19% to 21% of national health expenditures (NHE) during the induced demand phase. Cost savings materialize only after next-generation outcome-guaranteed payment models replace current FFS/MA paradigms or private-sector actors remove diagnosis from the reimbursement system entirely.

A critical additional finding: Administrative automation is the most likely bridge financing that can solve the cumulative break-even problem. Even under the most aggressive clinical AI displacement scenarios, cumulative NHE remains above baseline through 2040 because the induced demand overhang in the early years of adoption is too large. However, holding applicable administrative costs flat (at $317 billion) or reducing them 10% annually—which is applicable with current AI technology and independent of payment model reform—can pull the cumulative break-even forward to 2033 to 2038, making the overall AI investment fiscally viable within a 10-year CBO scoring window.

2. Methodology: The cognitive premium framework

We employ a subtractive methodology, starting from total NHE and removing every category involving physical intervention, tangible products, facility infrastructure, or nonphysician services to isolate the cognitive residual. We then apply a dynamic cognitive share that rises during induced demand and compresses during displacement. Administrative spend is modeled separately with three variants to capture AI automation effects independent of clinical AI adoption.

Exhibit 4. NHE scenario summaries

Metric	Value
Total NHE 2024	$5.3T
Cognitive premium 2024	$627B to $950B
Administrative spend 2024	Approximately $317B (6% of NHE)
Peak cognitive premium (est.)	$1.4T to 1.6T
Terminal cognitive share range	3% to 14%
Cumulative admin savings (flat through 2040)	Approximately $2.8T
Cumulative admin savings (-10%/yr through 2040)	Approximately $5T

Source: PitchBook • Geography: Global • As of March 31, 2026

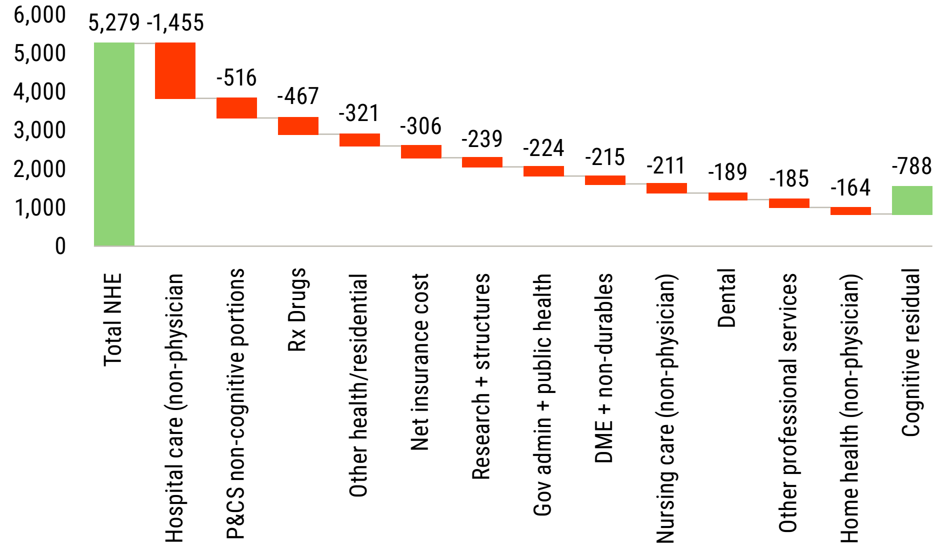
2.1 Data sources

- CMS National Health Expenditure Accounts (2024)—total and category-level spending ($5.28T)
- CMS Restructured BETOS Classification System (RBCS), 2023/2025—service-level decomposition of $1.14T in Medicare Part B allowed spending
- SOA Getzen Model of Long-Run Medical Cost Trends (2026–2036+ update)—baseline trend at ~5.0% nominal annual growth
- MedPAC physician fee schedule analyses; HHS OIG E&M reports—E&M spending shares and coding trends
- CBO (2008), Lancet Digital Health (2019), NBER WP 30857—technology-induced demand dynamics
- Oliver Wyman (March 2026)—Great Health Productivity Reset

2.2 The subtractive approach

Starting from $5.3 trillion, we subtract nonphysician hospital care (about $1,5 trillion), noncognitive physician and clinical services (around $516 billion), prescription drugs (about $467 billion), other health residential (about $321 billion), net insurance costs (around $306 billion), R&D and structures (about $239 billion), government administration and public health activities, durable medical equipment and nondurable medical products (about $215 billion), nonphysician nursing care (around $211 billion), dental (around $189 billion), other professional services (about $185 billion), and home health (around $164 billion). The cognitive residual is approximately $627 billion to $950 billion using ranges and approximately $790 billion at the midpoint.

Exhibit 5. Cognitive premium waterfall



Source: PitchBook • Geography: Global • As of March 31, 2026

2.3 The dynamic cognitive share

The cognitive premium share is not static. Every AI-surfaced diagnosis generates physician-dependent downstream activity: confirmatory E&M evaluation, treatment planning, specialist referrals, medication management visits, and professional interpretation of imaging/labs. During induced demand, the share rises (from 15% to 19% to 21%) because new utilization is disproportionately physician-dependent. During displacement, it compresses as AI absorbs cognitive work. This means the addressable market for AI disruption is $200 billion to $400 billion larger than static models project.

2.4 Administrative spend modeling

Administrative costs (approximately $317 billion in 2024, 6% of NHE) are modeled separately because AI automation of back-office functions is independent of payment model reform and clinical AI regulation. Three variants are applied across all five scenarios. Here we are conservatively assuming that just less than half (6%) of the low-end of common industry estimates of 15%-30% of NHE being administrative spend is applicable for reimbursement adjustments.

- **Grows with baseline:** Admin increases at the same rate as total NHE (default assumption in most healthcare cost models)
- **Held flat at 2024 level beginning in 2027 ($317 billion):** AI automation offsets organic growth—ambient scribes eliminate documentation labor, prior authorization is automated, claims adjudication becomes algorithmic, coding becomes AI-native
- **-10% per year:** Aggressive automation compounds annually beginning in 2027—the entire coding/billing/appeals/prior authorization apparatus is progressively replaced by AI systems

3. The three-phase AI impact model

Historical evidence shows new healthcare capabilities increase spending before efficiency gains materialize. The CBO documented that new diagnostics "invite much greater use and therefore tend to increase total spending."[51] The Lancet Digital Health noted that "introduction of low-price technologies might lead to an overall rise in expenditure" through addressing unmet demand.[52] Additionally, The Jevons Paradox states that, in the long term, an increase in resource-use efficiency will generate an increase in resource consumption rather than a decrease.

2036+ Phase 3 AI displacement	AI handles majority of cognitive work. Displacement dynamics depend on payment/ delivery model assumptions per scenario—ranging from modest reimbursement compression to diagnosis exiting the billing system entirely.
2031-2036 Phase 2 Regulatory catch-up	Regulatory frameworks mature. CMS recognizes AI-assisted care. Liability frameworks emerge. Next-gen payment models begin replacing FFS. Induced demand fades and costs converge to baseline.
2026-2031 Phase 1 Induced demand	AI diagnostic tools embed in primary-care workflows, patient-facing symptom checkers, and wearable-triggered screening. Marginal cost of diagnosis approaches zero. Under fee-for-service, millions of undiagnosed conditions surface, each generating downstream utilization. Under MA risk adjustment, AI supercharges HCC coding. The cognitive share rises to 19% to 21% as all new diagnoses require physician confirmation, treatment planning, and specialist referral.

4. Scenario definitions, assumptions, and parameters

Exhibit 6.

Parameter	Moderate	Aggressive	Bull case	Transformative	Trans.+ private
Phase 1 peak bump	+1.5%	+1.8%	+2.2%	+2.0%	+2.5%
Phase 2 end rate	-0.5%	-0.8%	-1.0%	-1.2%	-1.5%
Phase 3 mature rate	-2.5%	-4.0%	-5.5%	-6.0%	-7.0%
Displacement curve	Linear	Linear	Linear	Accelerating	Accelerating
Cog share: peak	18.5%	19.5%	20%	20%	21%
Cog share: P2 end	16%	15%	14%	13%	11%
Cog share: 2040	14%	11%	9%	5.5%	3%

Source: PitchBook • Geography: Global • As of March 31, 2026

4.1 Moderate scenario assumptions

Payment model: Current FFS and MA risk adjustment framework persists substantially intact. CMS makes incremental adjustments to the physician fee schedule and risk adjustment methodology but no structural reform. "Value-based care" remains primarily a risk-coding optimization exercise within Medicare Advantage.

Regulatory posture: FDA maintains conservative stance on autonomous AI clinical decision-making. AI tools approved as CDS requiring physician oversight. State medical boards require physician supervision.

AI adoption: Gradual penetration achieved via EHR-embedded decision support and ambient documentation, concentrated among large health systems. Small/independent practices lag.

Physician labor market: AMA and specialty society lobbying resists reimbursement compression. Residency slot caps not expanded. Scarcity premium remains intact.

Induced demand (+1.5%): Provider-initiated—patients screened during existing encounters using EHR-embedded tools. Modest additional utilization within the traditional visit-based model.

Exhibit 7. Six displacement mechanisms fire simultaneously:

Mechanism	Contribution	Description
Decision cascade savings	1.0–1.5%	Better decisions lead to fewer unnecessary referrals, optimized pathways
Next-generation payment transition	0.5–1.0%	Outcome-guaranteed models reward AI-delivered care
Chronic disease compounding	0.5–1.0%	Better management leads to lower acuity, which leads to deflationary utilization spiral
Physician labor repricing	0.3–0.5%	Scarcity premium erodes; compensation declines 20% to 30% real
Administrative collapse	0.3–0.5%	AI documentation, auto-adjudication, prior authorization elimination
Pharmacy optimization	0.2–0.4%	Right drug/dose/patient/time; fewer ADEs[D

Source: PitchBook • Geography: Global • As of March 31, 2026

Cognitive share (from 15% to 18.5% to 14%): Rises modestly with induced demand. Compresses only to 14% without structural payment reform because physician reimbursement rates decline slowly and the shift to supervisory roles is incomplete.

4.2 Aggressive scenario assumptions

Payment model: Next-generation outcome-guaranteed payment models emerge by the early 2030s. Key features: algorithmic risk benchmarking (replacing provider-documented HCC coding), multiyear accountability windows (three- to five-year contracts), hard outcome endpoints (hospitalizations, mortality, and functional status), technology-agnostic delivery (AI can deliver if outcome is met).

Regulatory posture: FDA creates streamlined pathway for autonomous AI in low-risk domains (chronic disease monitoring, medication titration, and screening interpretation) by 2032 to 2033. CMS permits AI-generated treatment plans to satisfy documentation requirements for specified conditions.

AI adoption: Broad-based by 2030 including independent practices via cloud platforms. Virtual-first care using AI triage captures 15% to 25% of primary-care encounters by 2035.

Physician labor market: Reimbursement per cognitive encounter declines 15% to 25% by 2040. Physicians shift to supervisory/exception-handling roles.

Cognitive share (from 15% to 19.5% to 11%): Higher peak from broader adoption. Deeper compression because outcome-guaranteed models incentivize AI-delivered services.

Critical assumption: Aggressive and above are conditional on structural payment reform. Current VBC/MA risk adjustment is a coding optimization game—part of the problem, not the solution. Without reform, moderate is the ceiling.

4.3 Bull case scenario assumptions

Payment model: Same outcome-guaranteed framework as aggressive but adopted faster. 60%+ covered lives under outcome-guaranteed arrangements by 2035. CMS implements algorithmic risk benchmarking for all of Medicare by 2033.

Cognitive share (from 15% to 20% to 9%): Highest nonprivate-sector peak. Compresses as all six mechanisms fire but diagnosis remains within the reimbursement system.

4.4 Transformative scenario assumptions

Fiscal forcing function: Federal health spending approaches $3 trillion by 2033. If AI adds 2% to the healthcare trend from 2026 to 2030, the federal share is approximately $20 billion to $30 billion/year in unbudgeted Medicare/Medicaid spending—$100 billion to $150 billion over five years, widening the deficit.

The payback math: Policymakers operating under CBO scoring windows and debt ceiling constraints need out-year savings to exceed near-term excess on a present-value basis:

1. CMS reprices cognitive services aggressively. AI provides political cover.
2. Congressional action on MA risk adjustment becomes unavoidable. AI-supercharged HCC coding inflating MA overpayments by $50 billion to $100 billion becomes a deficit target.
3. Phase 3 reimbursement compression is steeper and faster than organic displacement—policymakers cut first, delivery system adapts.

Displacement curve: Accelerating (quadratic) rather than linear—cuts come harder in 2038 to 2040 than 2036 to 2037.

Cognitive share (from 15% to 20% to 5.5%): May temporarily overshoot below long-run equilibrium (7% to 8%) as policymakers claw back Phase 1 deficit excess.

4.5 Transformative + private-sector scenario assumptions

The economic analogy: Google made search free (monetized via advertising). Robinhood made trading free (payment for order flow). GPS became free (location data). In each case cognitive function was commoditized to zero, utilization exploded, monetization shifted adjacently. Healthcare diagnosis is next.

Private-sector actors and incentives:

- **Amazon Health / One Medical:** Free AI triage leads to Amazon Pharmacy, Clinic, Prime Health. Diagnosis equals customer acquisition cost.
- **Apple/Google:** Continuous AI diagnosis via wearables becomes device feature. Monetized through hardware margins.
- **Retail health (CVS/Walgreens/Walmart):** Free AI screening leads to prescription customer conversion.
- **Health plans:** Under capitation, free AI diagnosis equals cost-avoidance investment.
- **Employer-direct:** Self-funded employers deploy free AI, bypassing insurance billing.

What makes trans.+ fundamentally different: Diagnosis stops being a revenue center and becomes more of a loss leader. The cognitive spend migrates from healthcare billing events to tech/retail P&Ls (device margins, pharmacy margins, health plan MLR).

Phase 1 is worst (+2.5%): Free consumer-facing diagnosis drives maximum utilization—no copay, no appointment needed. Cognitive share peaks at 21%.

The deficit hole is deepest, making fiscal clawback most aggressive. Unbudgeted excess of $100 billion to 175 billion.

Terminal cognitive share of 3% equals complex institutional care only (ICU, surgical pre-auth, tumor boards, psychiatric therapeutic relationship, supervisory sign-off).

5. Why current value-based care" Is not the mechanism

Current VBC models—particularly MA risk adjustment—are a risk-coding optimization game. MA plans code higher acuity than FFS for equivalent populations. Total Medicare spending on MA enrollees frequently exceeds FFS costs. AI is a coding tool in this model, not a care tool.

Meaningful displacement requires five structural features: (1) algorithmic risk benchmarking; (2) multiyear accountability; (3) hard outcome endpoints; (4) technology-agnostic delivery; (5) transparent real-time cost accounting. These define next-generation outcome-guaranteed payment models.

6. Administrative automation: The bridge financing

6.1 The cumulative break-even problem

Even under the most aggressive clinical scenarios (trans.+ private, -7.0% Phase 3 displacement), cumulative NHE remains above baseline through 2040. The induced demand phase front-loads $400 billion to $600 billion in excess spending that compounds forward, and the displacement phase—which only generates meaningful savings from 2036 onward—cannot fully offset this within the projection window.

This is not a modeling error. It is the central policy problem of AI in healthcare.

6.2 Why admin automation is the solution

Administrative automation is the one AI savings category that is completely independent of payment model reform, FDA approval for clinical AI, and physician labor repricing. It is back-office software. Prior authorization automation is already being mandated by CMS. Ambient scribes are eliminating documentation labor now. Claims adjudication is increasingly algorithmic. Revenue cycle management is being replaced by AI platforms.

Approximately $317 billion of administrative spend is almost entirely automatable by current AI technology—not future AI.

6.3 Three admin variants

Applicable admin grows with baseline (default): Administrative spend increases at the same rate as NHE.

Applicable admin held flat at $317 billion: AI automation offsets organic growth. Cumulative admin: approximately $5.4 trillion. Cumulative savings vs growing baseline: approximately $2.1 trillion.

Applicable admin reduced 10% per year: Aggressive but not unprecedented for technology-driven workforce displacement. Admin falls from $317 billion to approximately $37 billion by 2040. Cumulative admin: approximately $2.6 trillion. Cumulative savings: approximately $5 trillion.

6.4 Impact on cumulative break-even

Under the Trans.+ Private scenario:

Exhibit 8.

Admin variant	Cumulative break-even year	2040 cumulative NHE delta
Grows with baseline	>2040	Positive (above baseline)
Held flat	Around 2037 to2038	Modestly negative
-10% per year	Around 2031 to 2032	Substantially negative

Source: PitchBook • Geography: Global • As of March 31, 2026

The -10%/year admin variant pulls cumulative break-even to the Phase 1/Phase 2 boundary. Policymakers do not need to wait for Phase 3 displacement if admin automation generates $50 billion to $150 billion/year in savings starting in 2026 to 2027.

6.5 Second-order effect

Applicable admin is 6% of NHE. A 10% annual reduction in admin alone shaves approximately 0.6% off the total NHE growth rate. This compounds, making the overall cost trajectory more favorable before Phase 3 cognitive displacement kicks in.

Exhibit 9: Trend rates

Year	Baseline (5.0%)	Moderate	Aggressive	Bull Case	Transformative	Trans.+ Private
2024	5	5	5	5	5	5
2025	5	5	5	5	5	5
2026	5	5.3	5.4	5.4	5.4	5.5
2027	5	5.6	5.7	5.9	5.8	6
2028	5	5.9	6.1	6.3	6.2	6.5
2029	5	6.2	6.4	6.8	6.6	7
2030	5	6.5	6.8	7.2	7	7.5
2031	5	6.1	6.3	6.6	6.4	6.7
2032	5	5.7	5.8	5.9	5.7	5.9
2033	5	5.3	5.2	5.3	5.1	5.1
2034	5	4.9	4.7	4.6	4.4	4.3
2035	5	4.5	4.2	4	3.8	3.5

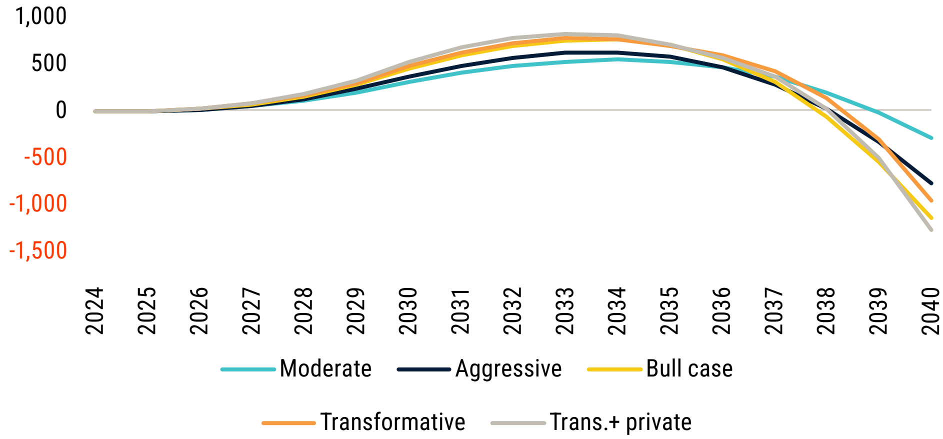
Source: PitchBook • Geography: Global • As of March 31, 2026

Exhibit 9: Trend rates (continued)

Year	Baseline (5.0%)	Moderate	Aggressive	Bull Case	Transformative	Trans.+ Private
2036	5	4.1	3.6	3.1	3.6	3.3
2037	5	3.7	2.9	2.2	3	2.6
2038	5	3.3	2.3	1.3	2.1	1.5
2039	5	2.9	1.6	0.4	0.7	0
2040	5	2.5	1	-0.5	-1	-2

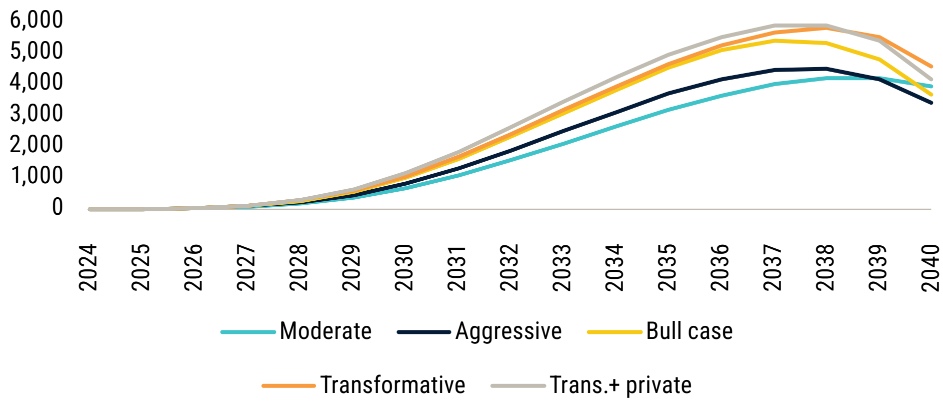
Source: PitchBook • Geography: Global • As of March 31, 2026

Exhibit 10: Annual NHE expenditure delta by scenario – administative expenses grow with medical trend



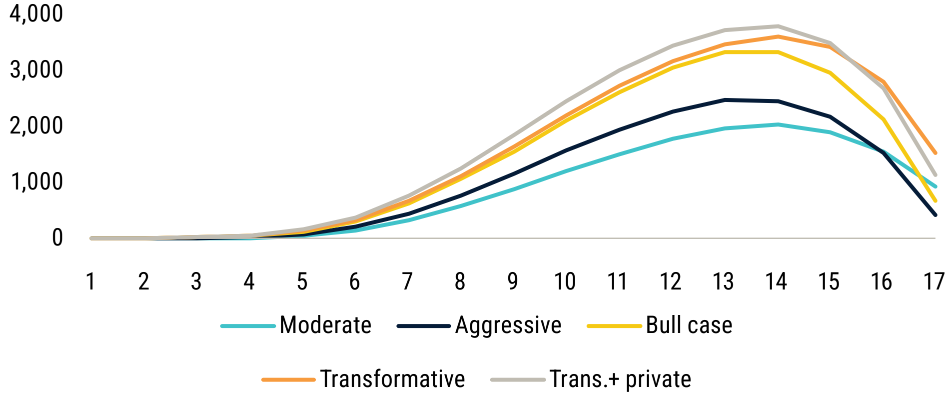
Source: PitchBook • Geography: Global • As of March 31, 2026

Exhibit 11: Cumulative NHE expenditure delta by scenario – administrative expenses grow with medical trend



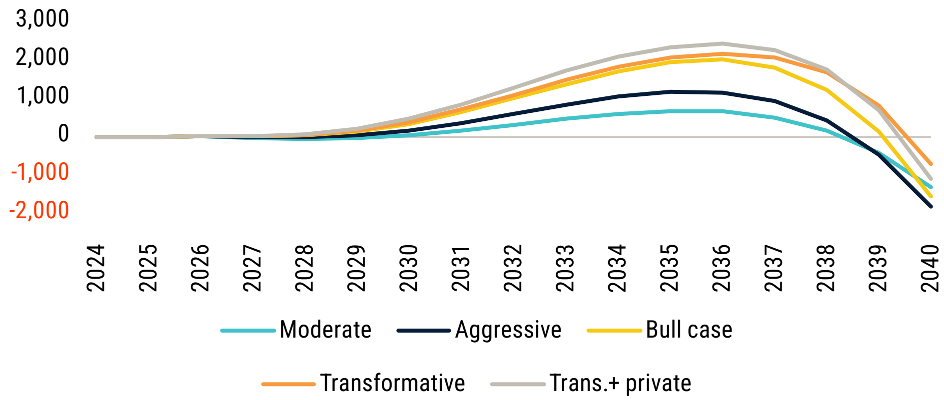
Source: PitchBook • Geography: Global • As of March 31, 2026

Exhibit 12: Cumulative NHE Expenditure Delta—Flat Admin Spend



Source: PitchBook • Geography: Global • As of March 31, 2026

Exhibit 13: Cumulative NHE Expenditure Delta—Admin Cut



Source: PitchBook • Geography: Global • As of March 31, 2026

See Interactive Model (for PitchBook subscribers only) for computed results across all scenario/admin combinations

Projection variance drivers

One thing is certain about the financial implications of AI in healthcare delivery: The models—including ours— are "wrong," but what are the most likely drivers of variance? We believe the magnitude of that inaccuracy will most likely be driven by the following dynamics: 1) 2025 and 2026 cost trends are likely higher than the Getzen model baseline of 5%, 2) the Getzen model and our adjustments do not make specific assumptions with regard to higher medication costs for personalized medications with curative potential, 3) similarly, neither Getzen's nor our model specifically contemplates the impact of breakthroughs in medical device technology such as Neuralinks' recent developments in allowing ALS patients to speak and to its goals of being able to bypass damaged neural pathways, 4) a recent study in the NEJM AI that indicated AI diagnosis can lower medical costs through downshifting medical site of care to the most appropriate setting, leading to greater primary-care practice setting utilization at the expense of more costly specialist care settings and hospital settings,[53] and 5) the potential reduction in medical errors (the Society of Actuaries put this cost at $19.5 billion annually in 2008, which would be $35 billion to $40 billion in 2026 adjusted for medical inflation or 60 basis points to 70 basis points of total NHE projections for 2026).[54]

This analysis does not factor in the impact of higher coding intensity that allows providers to generate higher compensation than prior to using AI in administrative functions. Our view is that the payers had a head start on AI use in claims denials and that the recent advancements in clinical documentation have leveled the playing field, allowing providers to receive appropriate compensation. As AI clinical documentation further penetrates the market, providers will see continued benefit through this adoption curve.

Additionally, other AI billing technologies are likely being deployed that are more questionable, but that is a subject for another day. We do not factor fraud into our analysis, as we believe fraud is generally a separate issue in terms of medical care delivery costs, and AI's use in care delivery does not materially change this dynamic. However, addressing healthcare fraud could provide a source of funding for higher near-term medical costs, although the political will to address this issue appears to be absent as there seem to be too many hands in the cookie jar.

When we compare the risks of these competing impacts between nearer-term higher costs and longer-term higher costs, they appear to us to be somewhat balanced. We believe it is most likely healthcare costs will go higher than baseline in the near term, as evidenced by higher-than-expected health insurance medical costs in 2025. This will likely outweigh long-term benefits without substantially improved health outcomes, which require behavioral changes that underwriting regulations currently restrict. Additionally, if near-term cost trends are higher than expected, longer-term cost improvement opportunities will likely be viewed more skeptically from an actuarial perspective.

One can calculate a wide range of the value of physician cognition, and our methodology assigns more value than most. The most logical areas to fund higher near-term costs are administrative savings, which are already being implemented successfully, although reimbursement levels are not factoring their benefit. It seems likely that physician cognition will likely be a long-term cost target as well, although these are policy choices, and other changes in reimbursement policies are possible. However, physician cognition seems more likely to be automated in the mid-term than surgeries. Our analysis is a framework for likely cost pressures and offering potential regulatory responses, none of which are simple or easy. However, the sooner we understand the likely implications, the sooner policy makers can alter course.

All projections in nominal dollars. Sensitive to assumptions about baseline trend, AI adoption velocity, regulatory response, payment model evolution, private-sector disruption, and administrative automation velocity.

References

1: "More Than 80% of Physicians Use AI Professionally: AMA Survey," AMA, Kevin B. O'Reilly, March 12, 2026.

2: ibid.

3: ibid.

4: ibid.

5: ibid.

6: ibid.

7: ibid.

8: "Many in US Consider AI-Generated Health Information Useful and Reliable," Annenberg, July 14, 2025.

9: ibid.

10: ibid.

11: ibid.

12: ibid.

13: "Introducing HealthBench," OpenAI, May 12, 2025.

14: "MedAgentBench: A Realistic Virtual EHR Environment to Benchmark Medical LLM Agents," Stanford ML Group, Yixing Jiang, et al., n.d., accessed April 10, 2026.

15: "Diagnostic Performances of GPT-4o, Claude 3 Opus, and Gemini 1.5 Pro in 'Diagnosis Please' Cases," Jpn J Radiol, Yuki Sonoda, et al., July 1, 2024.

16: "ChatGPT Health Performance in a Structured Test of Triage Recommendations," Nature Medicine, Ashwin Ramaswarmy, et al., February 23, 2026.

17: Dr. Ami Parekh, Chief Health Officer at Included Health, interviewed by author, March 24, 2026.

18: "Introducing ChatGPT Health," OpenAI, January 7, 2026.

19: "Doctronic Raises $40M Series B Following Breakthrough as First AI to Legally Renew Prescriptions in the US," Businesswire, March 24, 2026.

20: "Our Acquisition of Summa Health," General Catalyst, Hemant Taneja and Marc Harrison, January 17, 2024.

21: "Meet Your Personal Doctor," JIVI, 2026, accessed April 10, 2026.

22: "Vinod Khosla: AI Tutors, Doctors & Farmer Experts For 1.5B Indians," YouTube, uploaded by Business Today, 2026, accessed April 10, 2026.

23: "Agent Hospital: A Simulacrum of Hospital with Evolvable Medical Agents," arXiv, Junkai Li, et al., January 17, 2025.

24: "Tsinghua University Researchers Build Virtual Agent Hospital for AI Doctor Training Without Human Intervention," NotebookCheck, David Chien, June 16, 2024.

25: "Inside China's Agent Hospital: AI Doctors Get Residency," Medium, Celine Liu, May 26, 2025.

26: “Tsinghua University Launches AI-riven Hospital to Train Next-Gen Doctors,” TechNode, April 29, 2025.

27: “AI Agent Doctors Score 93% in Diagnostics at China’s Virtual Hospital, Surpassing Humans,” nurse.org, Angelina Walker, August 11, 2025.

28: “Large-Scale Local Deployment of DeepSeek-R1 in Pilot Hospitals in China: A Nationwide Cross-sectional Survey,” medRxiv, Meng Yuan, et al.,May 16, 2025.

29: “DeepSeek Deployed in 90 Chinese Tertiary Hospitals: How Artificial Intelligence Is Transforming Clinical Practice,” J Med Syst, Jishizhan Chen and Chunying Miao, April 24, 2025.

30: “DeepSeek’s AI in Hospitals Is ‘Too Fast, Too Soon’, Chinese Medical Researchers Warn,” South China Morning Post, Xinmei Shen, May 14, 2025.

31: “What if an FDA-Authorized Clinical Agentic AI Could Provide Safe and Effective Cardiovascular Care to Every American? ARPA-H, 2026, accessed April 10, 2026.

32: “ARPA-H to Revolutionize Cardiovascular Disease Management With Clinical Agentic AI, ARPA-H, January 13, 2026.

33: “NCQA’s Proposed Timeline for Retiring and Replacing HEDIS Hybrid Measures,” NCQA, November 15, 2024.

34: “ARPA-H Launches Exploration Topic to Improve Chatbots for Patient-Facing Applications,” ARPA-H, April 19, 2024.

35: “Stanford, Princeton Scientists Launch MedOS AI-XR-Cobot Clinical System,” The Robot Report, Eugene Demaitre, February 5, 2026.

36: “Sprinter Health Increases Healthcare Access Through Expansion of Preventive Care+ Visits Across 25 States,” PR Newswire, February 19, 2026.

37: “Sprinter Care+ Visits Strengthen Engagement With Primary Care,” Sprinter Health, 2026, accessed April 13, 2026.

38: “Rural Roots, National Impact: Inside Geisinger’s Innovation Engine, Spotify, uploaded by Verizon, March 2, 2026.

39: “ACCESS (Advancing Chronic Care With Effective, Scalable Solutions) Model,” CMS, 2026, accessed March 15, 2026.

40: “Aging Patients and Doctors Drive Nation’s Physician Shortage,” AAMC, Patrick Boyle, June 11, 2021.

41: “The Association between Health Professional Shortage Area (HPSA) Status, Work Environment, and Nurse Practitioner Burnout and Job Dissatisfaction,” J Health Care Poor Underserved,” Amelia E. Schlak, et al., July 22, 2022.

42: “A Giant Leap: How AI is Transforming Healthcare and What That Means for Our Future,” R.M. Wachter, Penguin Random House, 2026.

43: “Clinical Decision Support Software,” US FDA, January 2026.

44: “FDA “Cuts Red Tape” on Clinical Decision Support Software and Wearable Products for General Wellness,” Arnold & Potter, Abeba Habtemariam, et al., January 21, 2026.

45: Primary Care Health Professional Shortage Areas (HPSAs),” KFF, December 31, 2025.

46: “Healthcare on Air by Verizon,” Spotify, uploaded by Healthcare on Air by Verizon, 2026, accessed March 15, 2026.

47: Referenced work: "Health Workforce for Health Equity," Medical Care, Patricia Pittman, et. al., October 2021.

48: "DigitalHealth and the Next Chapter of Healthcare Decision Making," Spotify, uploaded by Healthcare on Air by Verizon, January 7, 2026.

49: "Rural Roots, National Impact: Inside Geisinger's Innovation Engine," Healthcare on Air by Verizon, March 2, 2026.

50: "A Giant Leap: How AI is Transforming Healthcare and What That Means for Our Future," R.M. Wachter, Penguin Random House, 2026.

51: Ibid.

52: "Technological Change and the Growth of Health Care Spending," CBO, January 2008.

53: "Digital Health and the Elusive Quest for Cost Savings," The Lancet, Kazem Rahimi, July 2019.

54: "From Advice to Action—Real-World Behavior of Patients Using an Integrated Diagnostic Decision Support System for Navigating the Health Care System," NEJM, Fabienne Cotte, et al., March 5, 2026.

55: "The Economic Measurement of Medical Errors," Society of Actuaries' Health Section, Jon Shreve, et al., June 2010.

PitchBook provides actionable insights across the global capital markets.

Additional research:

Healthtech AI Scribes

Download the report here

Healthcare services AI agents a $155 billion opportunity

Download the report here

PitchBook Insights is an online compendium of in-depth data, news, analysis, and perspectives that shape the private capital markets.

PitchBook subscribers enjoy exclusive access to a comprehensive suite of private market insights, including proprietary research, news, data, tools, and more on the PitchBook Platform.

Nizar Tarhuni
Executive Vice President of Research and Market Intelligence

Paul Condra
Senior Director, Global Head of Private Markets Research

James Ulan
Director, Industry & Technology Research

Report created by:

Brian Wright
Lead Analyst—Healthcare

Ben Zercher
Senior. Analyst—Biotech & Pharma

Ben Riccio
Analyst

Adi George
Associate Data Analyst

Adriana Hansen
Graphic Designer

Jenna O'Malley
Senior Graphic Designer

Learn more about PitchBook's Institutional Research team.

Click here for PitchBook's report methodologies.