o3 on AI and Health Spending

June 19, 2025

Below is a “first-principles” walk-through of how large-language-model (LLM) and broader AI tools are likely to affect U.S. national health expenditures (NHE). I lean on published facts only for baseline magnitudes (device approvals, spending totals, price points, etc.); the causal logic, scenarios, and net-impact estimates are my own.

---

1. Where the money is today

Spending component	2023 baseline	Share of NHE	Why it matters for AI
Total NHE	$4.9 T	100 %	Target to be “bent” by AI (cms.gov)
Administrative overhead (claims, billing, prior-auth, documentation, scheduling, supply chain)	≈30 % of NHE (≈$1.4 T) (commonwealthfund.org)	Biggest pool of non-clinical costs; largely information work	
Clinical delivery (hospitals, physician practices, imaging, labs, drugs)	≈60 %	Mix of labor-intensive and capital-intensive tasks; many AI touchpoints	
R&D and technology capital	≈10 % (includes drug discovery, IT, equipment)	AI both saves (faster R&D) and adds (new high-cost therapies)	

---

2. Direct cost-reducing vectors

Vector	Mechanism	Order-of-magnitude potential
Back-office automation (eligibility, claims edits, denials, scheduling)	LLM-powered copilots + RPA reduce keystrokes and re-work	13-25 % admin-cost cut estimated by McKinsey → $180-350 B/yr if fully realized (mckinsey.com, cms.gov)
Clinical documentation	AI scribes ($99-299 / mo vs $2,800 / mo for live scribes) free ~2 h/MD/day and save ≈$30 B/yr nationally (pmc.ncbi.nlm.nih.gov, getfreed.ai)	
Radiology & imaging	340 + FDA-cleared algorithms triage "normal" studies, cutting reading time 15-30 % and enabling partial task-sharing that models show can trim mammography program costs by up to 30 % (washingtonpost.com, medicine.illinois.edu)	
Robotic & AI-assisted surgery	Shorter LOS (-2 days), 50 % fewer readmissions; when scaled can offset high capital costs and deliver net savings in procedures with large volumes (sermo.com, facs.org)	
Remote patient monitoring & virtual wards	AI-driven alerts cut readmissions 20-45 %; each avoided readmission saves ≈$15 k (media.market.us, formative.jmir.org)	
Supply-chain and workforce optimisation	Predictive inventory and AI scheduling trim labour/waste 5-15 %; supply-chain is 39 % of patient-care cost base, so even a 10 % efficiency yields ≈$20 B/yr across hospitals (pwc.com, myshyft.com)	

Bottom-line potential, if everything above were fully adopted and efficiency passed through to payers: ≈$250-400 B a year (5-8 % of NHE) by early-2030s. That alone would slow the CMS-projected 5.6 % annual growth path to roughly 4.5 %. It is a meaningful "bend in the curve," but not an outright reversal.

---

3. Cost-increasing (and cost-shifting) vectors

Vector	Why it pushes spending up
Up-front implementation bills	Integrating AI into EHRs often costs $100 k – $500 k per site, plus recurring cloud/compute fees and regulatory compliance costs (callin.io)
Supply-induced demand	Easier documentation, chat triage and imaging can uncover previously unmet need and raise utilisation (classic "Jevons effect" seen after MRI and CT adoption).
AI-enabled high-complexity therapies	Precision oncology & bespoke cancer vaccines can exceed $100 k per course—even if discovered more cheaply with AI (businesswire.com, ft.com)
Algorithmic denial tools	Payer-side AI sometimes denies care 16× more often, shifting costs (and appeal work) back to providers and patients (ama-assn.org)
Cyber-risk & monitoring overhead	More software → larger attack surface → greater ongoing spend on security, audit, model validation.
Equity gap/duplication	Early-adopter hospitals buy AI twice—first bolt-on apps, then model-ready EHR upgrades—while late adopters keep legacy workflows, blunting system-wide savings.

---

4. Drug discovery: a special case

AI platforms (e.g., Insilico completing pre-clinical work in 18 mo for $2.6 M) slash R&D cycle times (insilico.com). Yet pharma pricing is value-based, not cost-plus; unless payers get stronger negotiating power, the lion's share of R&D savings will enlarge margins rather than lower list prices. In other words, AI may *raise* probability-weighted pipeline returns while doing little to ease downstream spending.

---

5. Short-, medium-, and long-run outlook

Horizon	Net pressure on NHE	Rationale

0-5 years (2025-29)	**Slightly inflationary**	Heavy capital outlays, hype-driven over-ordering of AI add-ons, and expensive new precision drugs hit before efficiency gains fully scale.
5-10 years	**Neutral to mildly deflationary** (-0.5 pp off annual growth)	Back-office automation, AI scribes, radiology task-sharing and supply-chain AI mature; but induced utilisation and high-cost therapeutics offset some wins.
10 + years	**Depends on policy**	If savings are reinvested in population health or captured via payment reform, AI could meaningfully flatten the cost curve. If instead the system simply layers AI on top of today's fee-for-service incentives, total spend may still rise, albeit from a higher-quality baseline.

---

6. Strategic implications for payers, providers and policymakers

1. **Target the fat first.** Administrative and documentation automation gives the fastest, lowest-risk ROI; make those savings a funding source for more experimental clinical AI.

2. **Use payment design to keep savings.** Bundled payments or global budgets ensure that efficiency doesn't just expand volume.

3. **Invest in evaluation infrastructure.** Real-world cost-effectiveness registries for AI tools (analogous to drug HTA) will be critical.

4. **Address denial-of-care optics early.** Transparent audit trails and appeal protections are needed before payer-side GPTs trigger a backlash.

5. **Reskill, don't just downsize.** AI that frees clinicians from clerical work but leaves workforce ratios unchanged simply raises quality; to lower cost you must redesign care pathways around new task division.

Bottom line

LLM-based and other AI systems are poised to shave tens of billions off administrative waste and selected clinical workflows—but they will simultaneously unleash new, often expensive capabilities that expand what medicine can do. Absent accompanying payment reform, the most

plausible outcome is **slower growth rather than an absolute drop** in national health spending. Whether AI ultimately bends the cost curve *enough* therefore hinges less on the algorithms themselves than on the policy and market structures into which they are deployed.