



First to the Floor

Why Operating History, Not Technology,
Will Determine the Winners in Physical AI

Roger Carter

VP of Engineering, Prime Movers Lab

March 2026



The cheaper robots get, the more valuable their operating history becomes. This is the central paradox of the humanoid robotics economy. Investors chase the hardware cost curve or the next foundation model benchmark, assuming that whoever builds the best robot wins. That assumption has a shelf life. Hardware costs have fallen 40% in a single year and modern AI models are commoditizing the intelligence layer. Two years ago, teaching a robot a new task required months of custom programming. Today, foundation models trained on video can get a robot to useful performance in hours. In 36 months, the early advantages in this market will be very difficult to reverse. The differentiator will not be who built the best robot. It will be who put it to work first: who logged the failures, survived the edge cases, and accumulated the actuarial record that lets an underwriter price a policy and a CFO sign a purchase order. The moat is not technical. It is historical.

150,000+

Warehouses and factories worldwide designed for human workers

80%

Of existing industrial facilities still operate without automation

2.1M

Projected manufacturing job gap in the US by 2030

Why Humanoids, Why Now

Over 150,000 warehouses and factories worldwide were designed for human workers: human-width aisles, human-height shelves, human-operated tools. Retrofitting these spaces for traditional automation is prohibitively expensive. Purpose-built robots excel in new facilities engineered around their capabilities. However, this is a fraction of the installed base. 80% of existing industrial facilities still operate without automation and are designed around human labor. A humanoid form factor promises to operate in these environments without requiring custom conveyor integration or work space engineering that multiplies the hardware cost of a traditional industrial arm over its lifetime.

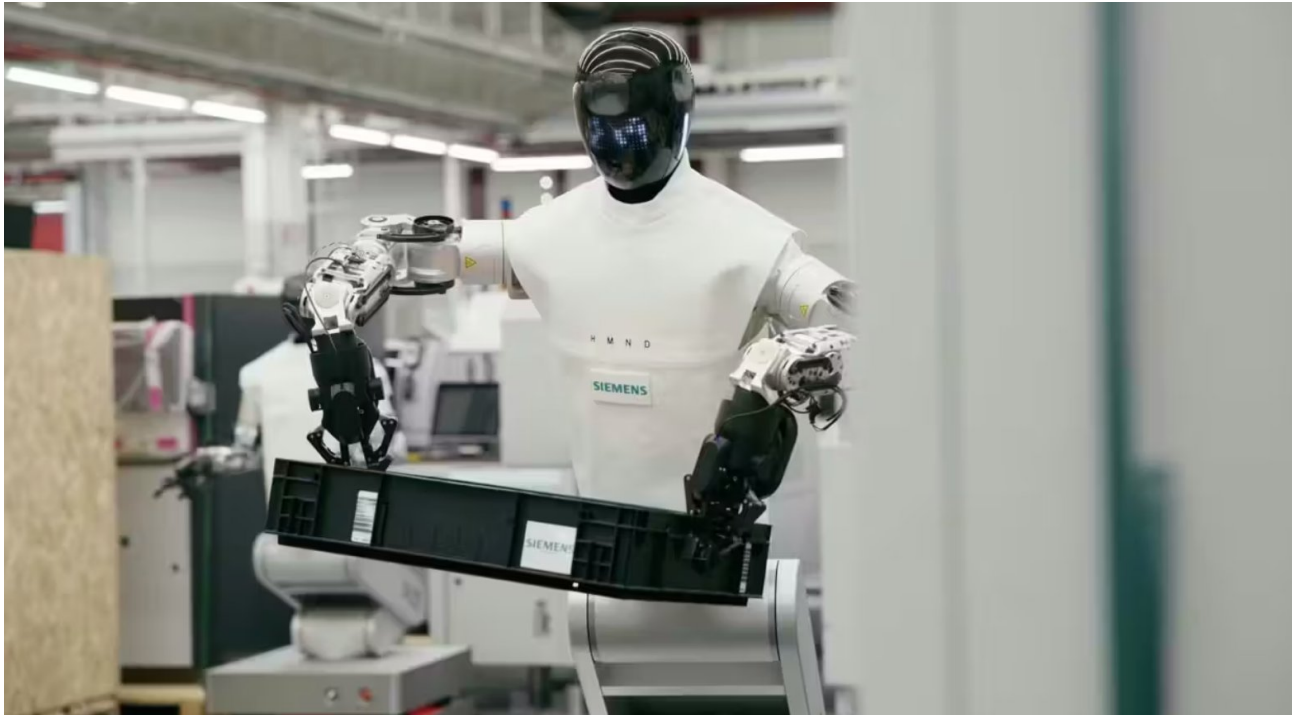
The differentiator will not be who built the best robot. It will be who put it to work first: who logged the failures, survived the edge cases, and accumulated the actuarial record that lets an underwriter price a policy and a CFO sign a purchase order. The moat is not technical. It is historical.

The World Economic Forum found that 63% of manufacturers say the primary barrier to automation isn't cost or technology, it's finding people who can run it. Instead of requiring highly skilled programmers, general purpose humanoid robots offer the promise of being trainable by the average worker: show the robot the motion, correct it in plain language, assign the next task, repeat. The technical onboarding barrier the WEF is measuring is not workforce development at scale, it's the cost and complexity of telling a robot what to do.

The barrier is found not just in hiring workers who can run these robots; it's found in hiring workers at all. The United States had approximately half a million unfilled manufacturing openings as of

January 2026. Deloitte projects over two million unfilled manufacturing jobs by 2030 with an estimated trillion dollar annual cost. Japan, where the working-age population has fallen by four million in a decade and a third of citizens are over 65, set all-time records for robotics orders in Q1 2025.

This is not enthusiasm for new technology, it's a well developed economy employing robotics because it has exhausted every other option. The US and Europe are on similar demographic curves, albeit earlier in the progression. Japan, Germany, and China face demographic trajectories that no combination of immigration, wage policy, or workforce training is projected to reverse at scale. The question is no longer whether automation fills these roles, it's which platform fills them first.



The HMND-01 Alpha operating in a Siemens manufacturing environment.

Which Humanoids

Choice of form factor determines how quickly a robot platform can begin accumulating the operating history necessary to gain market dominance. In Europe, wheeled humanoid platforms can certify today under the ISO standard 3691-4 (the established safety standard for autonomous mobile robots), meaning these robots can be deployed into production environments today, uncaged, to work alongside humans. Bipedal platforms must wait for ISO 25785-1, a humanoid-specific standard that is unlikely to reach Draft International Standard before 2029, or be enclosed in safety equipment that isolates them from working directly with humans.

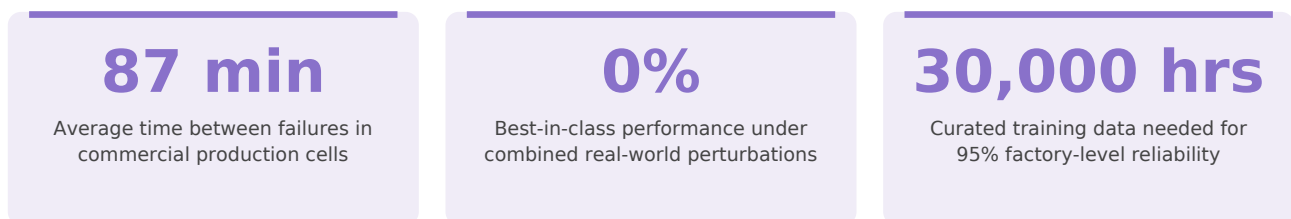
This certification advantage held by wheeled humanoids is not EU-specific. In the US, wheeled platforms can certify under ANSI B56.5 and UL 3100 today, and no published standard addresses bipedal humanoids in shared workspaces. McKinsey, Bain, and Bank of America all identify wheeled

mobile manipulation as the near-term commercial opportunity through 2028. In a race measured in deployed hours, the platforms that can legally operate in EU and US facilities now hold a structural head start over those waiting for new regulatory frameworks to be written.

87 Minutes

Why can't a well-funded latecomer just buy better AI and skip the operating-history phase? Because the reliability problem facing deployed robots has two layers, and only one of them responds to better AI.

The first layer is compound failure. A robotic system achieving 99% per-step reliability sounds high until you consider that an automotive assembly station runs 40 to 100 steps per cycle. On a 50-step workflow, 99% per-step success compounds to a 61% success rate on the process. At the 97% success rate that the best current models achieve on standard benchmarks, this 50 step workflow succeeds 22% of the time. In published stress tests, commercial production cells average a failure every 87 minutes meaning human intervention at least five times per shift. This layer has a known fix. Based on 131 published studies, reaching 90% factory-level reliability requires approximately 2,400 hours of curated training data. Reaching 95% requires approximately 30,000. Open-source foundation models will likely hit those thresholds within two years. That part of the advantage commoditizes.



The second layer does not. When the only variable changing is lighting between shifts, or parts arriving at slightly different angles, more training data closes the gap. When lighting, position, and layout all change at once, which they do on every factory floor every shift, today's best models' success rate collapses to zero. A February 2026 five-model benchmark confirmed the pattern across the field: best-in-class performance reached only 18-24% when multiple variables shifted simultaneously.

The reason is architectural. Current vision-language-action models are pattern-recognition systems. They identify visual sequences they have seen before and replay the associated motor commands; none maintain an explicit model of physics. A robot trained on ten thousand successful box transfers doesn't know that a wet conveyor belt changes the grip force required, or that a box three centimeters off-center torques the wrist differently. It knows what success looks like, and it doesn't know why it succeeds. More training data cannot close a functional gap that exists because the system has no causal model of the world it operates in. Closing that gap requires the robot to fail in a specific environment, be corrected, and accumulate enough corrections to generalize. That process happens on real factory floors, not in simulation.

\$7.55 an Hour

The business case for a humanoid robot turns on one question: can it do the job well enough that its amortized cost per hour beats the fully loaded cost of the person it replaces? Goldman Sachs documented a 40% single-year decline in humanoid manufacturing costs between 2022 and 2023. The trajectory follows Wright's Law, and the critical threshold arrives around mid-2027 when the bill of materials crosses approximately \$50,000.

At that price point, in 16-hour daily operation, the all-in effective cost falls to approximately \$7.55 per robot-hour on a fully loaded total cost of ownership basis. Today, sixteen hours requires the robot to either swap its own battery from a rack when it runs low, or stand on a charging station. Some platforms already support this but most bipedals don't yet. At current runtimes of 6-10 productive hours, the effective cost would be \$12-\$20/hour.

Use Case	Human Cost (Fully Loaded)	Robot at \$150K BOM	Robot at \$50K BOM	Closes Today?
Automotive assembly (US UAW)	\$65-\$85/hr	\$21.68/hr	\$7.55/hr	Yes
Automotive assembly (Germany)	\$32-\$39/hr	\$21.68/hr	\$7.55/hr	Yes
Coastal warehouse picking (2-shift)	\$35-\$47/hr	\$21.68/hr	\$7.55/hr	Yes
Food processing / meat packing	\$28-\$38/hr	\$21.68/hr	\$7.55/hr	Marginal
Rural warehouse picking	\$22-\$29/hr	\$21.68/hr	\$7.55/hr	Not yet
Elder care (Japan)	\$12-\$17/hr	\$21.68/hr	\$7.55/hr	Not until ~2029

Robot-hour costs reflect PML's internal TCO model: BOM amortized over 60 months at 16-hour daily utilization (5,840 hrs/yr) with 15% loading for maintenance, power, and insurance. BOM price points sourced via Goldman Sachs.

The declining hardware cost is also partially offset by a cost category with no precedent in traditional robotics: software. VLA model access, fleet management, enterprise ERP connectors, and task-specific fine-tuning add an estimated \$18,000-\$87,000 per robot annually. The range reflects the difference between a single-task warehouse deployment and a multi-task, ERP-integrated manufacturing cell. Either way, it's a line item that most published break-even analyses currently omit. (PML estimate; market pricing is early-stage and not yet publicly benchmarked.)

The economics are sector-dependent. At \$22/hour rural warehouse labor, the case does not yet close. At \$35/hour urban manufacturing on a single shift, it is marginal. At \$48/hour automotive OEM labor, it is already compelling, which is why automotive is where serious deployments are concentrating. At those rates, the robot pays for itself well inside the 18-month payback threshold that Goldman Sachs identifies as the institutional approval benchmark for capital equipment. The decision shifts from 'can we justify this' to 'can we afford to wait'. Tesla, which operates some of the

highest-volume automotive assembly lines in the world, is investing billions into Optimus. They see the labor math from the inside.

Cost parity is necessary but not sufficient. In the marginal sectors, the economics close on a spreadsheet before the technology closes on the factory floor. Food processing demands cold-chain tolerance, variable product geometry, and wet-surface handling that no current platform has demonstrated at production reliability. The sectors where deployment is actually concentrating are the ones where both the cost math and the technical capability overlap today. As reliability improves and operating hours accumulate, the overlap expands. Sectors that are marginal now become viable not because the hardware gets cheaper, but because the robots get better at working in those specific environments.

Each BOM threshold crossing opens a new tier of the addressable market, but the use cases that flip first are not the highest-wage; they're the highest-turnover. In sectors running 100–150% annual workforce cycling, recruitment, onboarding, and lost-productivity costs compound on top of base wages: a hidden multiplier most break-even models undercount. When a warehouse workforce turns over at 150% annually, turnover alone becomes a line item measured in billions. That's not a labor problem. It's a business model waiting for a different solution.

The Drawbridge

The operating-history moat doesn't just reward the early. It punishes the late. Three compounding dynamics create a deployment window that does not reopen for latecomers.

No data, no policy.

No insurer in the US or EU currently offers actuarially priced comprehensive humanoid coverage. Early-market estimates range from \$40,000–\$80,000 per robot per year: novelty pricing in the absence of data, not risk-priced coverage. At those rates, insurance alone could represent 10–25% of annual operating cost. Chinese insurer CPIC launched the world's first dedicated humanoid policy in September 2025 at ~\$700 per robot per year, a roughly 57–114x discount. That gap explains much of China's deployment speed advantage. It will persist until western insurers have enough real-world data to price humanoid coverage on the same basis, but it won't last forever. AgiBot opened a Munich office in February 2026; Chinese manufacturers establishing Western subsidiaries generate locally-domiciled incident data from day one, and that data qualifies. The moat this paper describes is real. The question is whether the operating-history leaders reach actuarial scale before Chinese manufacturers reach Western certification.

Regulation rewards the early.

Three major EU regulatory regimes take effect in the same eighteen-month window as the cost inflection: the Product Liability Directive extending strict liability to AI (December 2026), the Machinery Regulation's AI compliance provisions (January 2027), and the AI Act's high-risk obligations (December 2027, per the Digital Omnibus revision adopted by EU Parliament on March

26, 2026, pending final trilogue adoption). Companies deploying before these dates with documented safety protocols hold a fundamentally different legal position than those entering after. Once the Machinery Regulation takes effect, any over-the-air update to a high-risk AI system triggers a conformity assessment review. Not full re-certification in every case, but a documented process that consumer tech simply does not face. This creates a perverse incentive: fixing a safety-critical AI behavior may legally require full re-certification before deployment, rewarding companies whose architecture was designed for this constraint from the outset.

Standards cannot be written without data.

ISO 25785-1, the first humanoid-specific safety standard, cannot be finalized until sufficient real-world data on failure modes and incident rates exists to define quantitative safety thresholds. ISO TC 299 Working Group 12 has adopted a data-first methodology. The companies accumulating production data now are not merely preparing for the standard, they're helping write it. Contributors of real-world deployment data have outsized influence on how safety thresholds are defined, not because the process is captured, but because the data they provide is the only data that exists. In the US, the dynamic is even more direct. The White House AI policy framework, released March 2026, explicitly recommends against creating new federal rulemaking bodies and defers to industry-led standards.

What Compounds

Every hour of production deployment generates four compounding assets: training data that improves task performance, failure logs that inform safety protocols, customer-specific integrations that create switching costs, and actuarial loss history that eventually enables insurance pricing at scale. The infrastructure underlying those hours compounds in its own way. NVIDIA's GR00T platform and Jetson Thor architecture are becoming the default infrastructure for training and deploying embodied AI. Companies building within that ecosystem inherit 4.5 million CUDA developers and the most mature simulation-to-reality pipeline in the industry. Companies building outside it aren't competing against a chip. They're competing against a decade of tooling development that no one has a reason to leave. The concentration risk is real: a platform built within one ecosystem inherits its constraints as well as its advantages. The bet is that NVIDIA's lead in simulation-to-reality fidelity is durable enough to make that a worthwhile trade. The data layer is what makes it so.

Critics who argue that open-source foundation models will commoditize the AI training advantage are correct. They are wrong about the other three: failure mode documentation, customer integration depth, and actuarial loss history. A foundation model trained on internet video cannot generate the ISO-compliant, insurer-accepted, enterprise-referenceable incident record that a deployed production fleet accumulates over years of operation.

The commercial mechanism that makes early accumulation possible is Robotics-as-a-Service. The unit economics of humanoid deployment do not currently support outright purchase by most industrial buyers. A \$150,000 robot with an uncertain 30-month payback does not survive most capital allocation committees. Robotics-as-a-Service converts that conversation: a monthly contract

priced below the labor cost it replaces is an operating expense decision, not a capital one, and it clears a fundamentally different approval threshold. The consequence for the moat argument is structural. RaaS aligns the vendor's incentive with keeping the robot running, not just selling more of them. That means the companies deploying under this model are accumulating exactly the kind of verified, high-uptime production hours that insurers and procurement committees eventually need to see. The asset being built is not just the data. It is the track record, denominated in verified hours, that no later entrant can purchase.

Consider how Intuitive Surgical built the surgical robotics category. The company behind the da Vinci system went public in 2000 with \$26.6 million in revenue and now approaches \$200 billion in market capitalization. Its durability as a category leader over the following 25 years rests not primarily on technical differentiation; competitors have always existed. The compounding effect of surgical outcome data, trained surgeon networks, and FDA compliance history is what proved irreplicable.

The da Vinci robot was never a lasting moat. The twenty-five years of surgical outcome data was. Physical AI is heading for the same structure. The robot is not the moat. The operating history is.

Hours Logged ≠ Hours That Count

The most obvious challenge to this thesis is volume, and volume points to China. Chinese companies shipped 10–20x more humanoid units than Western competitors in 2025. If operating hours are the moat, does China already own it?

The distinction that matters is between hours that generate loss history that Western insurers and regulators can use, and hours that do not. Approximately 75% of Chinese shipments went to universities and research institutions, not commercial end-users. Educational and demonstration deployments do not generate the documented incident data, insurance claims history, or enterprise reference accounts that unlock Fortune 500 procurement committees.

The infrastructure gap reinforces this. China's cross-border data transfer framework classifies industrial incident data as "important data" subject to security assessment before transfer to foreign insurers or regulators. The autonomous vehicle precedent confirms regulators do not credit foreign miles for domestic approvals: Baidu's 240 million kilometers in China have not accelerated any US deployment approval.

This distinction is real but not permanent; the first moves are already visible. Chinese manufacturers establishing Western subsidiaries can generate locally domiciled hours from day one. The solar panel precedent is instructive: between 2009 and 2014, Western solar manufacturers held regulatory advantages over Chinese competitors. None survived quality parity. Robotics is structurally different. Solar panels are passive products with no ongoing incident history. A humanoid integrated into enterprise systems generates continuous liability data and creates switching costs with every deployment. But the lesson holds. The window is finite. By 2028, the moat must stand on its own merits.



The HMND-01 Alpha humanoid robot platform.

Proof on the Line

1,250 hrs

Figure AI's cumulative hours on
BMW's live production line

100,000+

Totes moved by Agility's Digit in
commercial warehouse operations

\$2.4B ARR

Humanoid's pre-orders from
Schaeffler, Ford, Siemens, Volvo,
IKEA

1,250 hours at BMW. Figure AI's deployment at BMW Spartanburg represents the most publicly documented automotive integration of a humanoid robot to date: approximately 30,000 vehicles produced with robotic assistance, 90,000+ parts handled, 1,250 cumulative hours on a live production line. Figure 02 loaded sheet-metal parts into welding jigs with 5-millimeter precision in 2-second cycles, running 10-hour shifts for eleven months. The forearm, the highest-failure-rate component, was completely re-designed for Figure 03 based on failure modes observed during Spartanburg.

BMW's decision to launch its first European humanoid pilot at Plant Leipzig, announced February 2026 with a different platform, reflects the same operating logic: documented production data, not a sales process, drives expansion decisions at this scale.

100,000 totes at GXO. Agility Robotics' Digit has moved more than 100,000 totes in commercial warehouse operations with GXO Logistics. The Mercado Libre agreement, with deployment beginning in San Antonio and expansion rights into Latin America, extends Agility's footprint beyond contract logistics into e-commerce fulfillment. Digit v5's 10:1 operating-to-charging ratio, versus the prior generation's 2:1, is a step-change in productive hours per shift. Agility is the only humanoid company with publicly confirmed commercial revenue across multiple Fortune-scale enterprise customers as of March 2026.

The Narrow Path. The third pattern emerging in early deployments is deliberate architectural restraint: teams that have watched what fails at scale and designed around it. The highest-profile humanoid programs have spent billions pursuing bipedal locomotion, custom actuators, and frontier AI simultaneously. The evidence points to a narrower path: deploy on a wheeled platform certifiable under existing AMR standards, build within the NVIDIA GR00T ecosystem rather than around it, and integrate directly into enterprise ERP systems so that every deployed hour generates structured, auditable data from day one.

Humanoid is the most direct current test of this thesis. Its leadership built Atlas at Boston Dynamics, scaled Brain Corp's fleet to 40,000 deployed robots, and ran perception for Yandex's autonomous vehicles. The HMND 01 is among the first robots designed to operate as a native node within SAP's Embodied AI agent layer, meaning it operates within the enterprise software infrastructure that 400,000 companies already run. Its deployment history is earlier-stage than Figure's or Agility's: proof-of-concept engagements at Siemens, Ford, and Schaeffler, not sustained production runs. If certifiability, deployment velocity, and enterprise integration depth compound faster than bipedal capability, the operating-history moat is already built by the time bipedal platforms are certified. The next two years of deployment data will make that visible.

What Would Change Our Minds

Any thesis worth holding is worth stress-testing.

A serious injury or fatality attributed to a deployed commercial humanoid triggers a regulatory enforcement action or enterprise-wide deployment freeze before any company reaches 500,000 cumulative operating hours. Between 1992 and 2017, NIOSH documented 41 robot-related workplace fatalities in the US, all involving traditional industrial arms rather than humanoids. The humanoid safety record is currently clean, but the deployed base is also negligible. As units in production environments scale from dozens to thousands, the probability of a serious incident scales with them. The relevant precedent is General Motors' Cruise program: a single pedestrian-dragging incident in October 2023 resulted in GM writing off \$1.9 billion, suspending all Cruise operations, and triggering NHTSA and California DMV investigations that effectively ended the program. A humanoid causing a serious injury on a production line that had been publicly promoted as safe could freeze Fortune 500 procurement decisions for 18 to 24 months. That does not invalidate the thesis, it compresses the time available to build the moat before conditions reset. A freeze delays every company equally, but the companies that have already accumulated production data, insurance relationships, and standards committee presence resume first. The companies that haven't deployed

yet are still at zero when the freeze lifts.

A foundation model achieves greater than 80% on LIBERO-PRO combined perturbations.

This would demonstrate that a robot can genuinely adapt to new situations rather than replaying memorized routines. Current state of the art: zero percent. Emerging work on world models that learn physics from raw video suggests a credible path, but no system has demonstrated this on real robot hardware yet. If crossed before January 2028, the AI training dimension of the moat dissolves before it can be built.

A Western insurer launches actuarially priced humanoid coverage without requiring minimum operating-hour history.

Current novelty pricing runs \$40,000–\$80,000 per robot per year; comparable industrial equipment is insured at \$750–\$3,000. A product priced below \$9,000 without a deployment-history requirement would signal that modeled data has substituted for observed data.

A sub-\$20,000 humanoid achieves commercial deployment at scale in Western markets

before any operating-history holder reaches the one-million-hour actuarial threshold. The test: 500+ units across 10+ enterprise sites with CE marking and disclosed insurance by end of 2027. If that happens, hardware cost was the binding constraint all along, and the operating-history thesis needs material revision.

The Floor is Open

The demographic imperative is irreversible. The cost curves are confirmed. The regulatory window is crystallizing in ways that will determine category leaders within 36 months. Commoditization does not destroy this thesis; it is the thesis. The \$50,000 inflection is where the economics first flip broadly. The longer-term trajectory toward \$25,000 is where hardware commoditizes entirely, concentrating all remaining value in the non-commodity layers above it: safety certification, actuarial data, and enterprise integration depth.

The companies that survive the pre-profitability deployment phase long enough to build that record will hold advantages that no amount of later capital can replicate, because time, by definition, cannot be purchased. Capital is necessary, but the race belongs to whoever converts it into deployed hours fastest. The clock is already running.