

Physical AI Is the Next Platform Shift, and Chat Was Only the On-Ramp

Basil C. Puglisi, MPA · basilpuglisi.com · October 2026

Capital and talent still flow to language models, and a growing share now follows those models toward the one place a chat window cannot reach, which is the physical world. Chat interfaces made AI legible to every board. Robots that see, plan, and act make AI consequential in warehouses, labs, hospitals, and homes, where a wrong output moves mass instead of words. Reading that shift as a chatbot with arms misses what changes, because the product surface changes, the data economy changes, and the cost of a model error changes with them.

The claim is not that deployment economics have already caught up with the technical frontier. They have not. The claim is that the model, simulation, data, and control stack is reorganizing around physical action fast enough to create a new platform layer. The gap between what a robot can do in a demonstration and what an operator can afford to run in production remains the test that matters. Tesla now supplies the clearest live case of that gap, because its humanoid robot and its driverless vehicles draw on shared training compute and leave a single public record (Tesla, 2026b).

This analysis follows an evidentiary discipline in which every fact must lead to a tactic, and every tactic must leave evidence, which the author calls Factics. The method was first published in 2012 and developed through the Digital Factics series (Puglisi, 2012, 2026c). Physical AI clears that bar only when teams stop grading demonstrations on how fluent the model sounds. Teams have to grade closed-loop work on in-spec task completion, recovery after failure, human intervention, and total cost in the real world.

Why are frontier investors treating embodiment as a new scaling axis?

Oliver Hsu, a partner on the American Dynamism team at Andreessen Horowitz, starts his April 2026 essay from the incumbent paradigm. Production AI is organized around language and code, and its scaling laws are well characterized (Hsu, 2026b). Its commercial flywheel of data, compute, and algorithmic improvement still pays (Hsu, 2026b). Hsu then names three adjacent domains leaving their gestation phase: robot learning, autonomous science in materials and

life sciences, and new human-machine interfaces from AR glasses to brain-computer interfaces (Hsu, 2026b). His argument holds that the largest gap between perceived capability and medium-term upside sits one step removed from the dominant paradigm (Hsu, 2026b). Those fields sit close enough to inherit its infrastructure and research momentum, yet far enough to demand nontrivial new work. That distance can become a moat against fast followers.

Hsu identifies five primitives the three domains share: learned representations of physical dynamics, architectures for embodied action, simulation and synthetic data infrastructure, an expanding sensory manifold, and closed-loop agentic orchestration (Hsu, 2026b). He treats the domains as mutually reinforcing. Robotics supplies the manipulation that self-driving laboratories need, and laboratories return physically grounded experimental data that can improve world models. Wearables and neural interfaces open new data channels for training embodied systems (Hsu, 2026b). That flywheel is the strategic object, because chat funded much of the stack and embodiment is where the stack finally meets matter.

The thesis deserves weight, and it still has to be read as what it is. A fund essay is an investment argument, and the firm's disclosure covers information that can come from portfolio companies (Hsu, 2026b). Hsu also published a sharper warning three months earlier. In January 2026, he described a wide gap between the robotics research frontier and production settings, where most robots remain narrowly preprogrammed (Hsu, 2026a). Research demonstrations of open-ended manipulation have yet to be reliably deployed at scale, and most humanoid deployments remain in pilot phases (Hsu, 2026a). Read together, the two essays make a stronger argument than either one alone, since the technical platform is forming faster than the deployment ledger.

That distinction matters because a platform thesis can be directionally right while an individual deployment is economically wrong. Anthropic's September 2026 robot exposure index estimates that present-day robots can perform 74 percent of physical work tasks in at least some settings (Legate-Yang & Massenkoff, 2026). The same study finds robots cost-competitive with human labor for only 0.3 percent of job tasks today (Legate-Yang & Massenkoff, 2026). The measure is not a universal factory cost study, but it supplies an independent economic boundary that vendor demonstrations do not.

The fact is that frontier capital analysis now frames robot learning, autonomous science, and new interfaces as compounding domains beyond language and code (Hsu, 2026b). Hsu's own deployment essay and independent economic evidence both show that deployment reliability and cost still trail technical reach (Hsu,

2026a; Legate-Yang & Massenkoff, 2026). The tactic is to redraw the AI portfolio map in three columns, digital agents, physical closed loop, and human interface expansion, then require every funded bet to state the operational variable it should move. The measure is the share of approved bets with a preregistered outcome, a frozen baseline, and a falsifier before funding begins, with a target of 100 percent.

What do GR00T, Gemini Robotics, and π_0 change about the build path?

NVIDIA introduced Project GR00T in March 2024 (NVIDIA, 2024). It then announced Isaac GR00T N1 in March 2025 as an open, fully customizable foundation model for generalized humanoid reasoning and skills (NVIDIA, 2025). The project lineage begins in 2024, while the N1 model cited here arrives in 2025. GR00T N1's dual-system design pairs a slower vision-language model that reasons and plans with a faster action model that turns those plans into continuous motion (NVIDIA, 2025). NVIDIA's technical write-up describes a training data pyramid, with internet-scale video at the base, synthetic data from the Omniverse platform in the middle, and real robot teleoperation at the peak (Vadrevu & Omotuyi, 2025). The same write-up reports a 40 percent improvement from adding synthetic data to real data in NVIDIA's own evaluation (Vadrevu & Omotuyi, 2025). That result is vendor-selected, so it belongs to the conditions of that evaluation rather than standing as a universal robotics gain. NVIDIA's announcement also named a collaboration with Google DeepMind and Disney Research on Newton, an open-source physics engine built for robot learning (NVIDIA, 2025).

Google DeepMind approached the same frontier from another direction that month. Its announcement described Gemini Robotics as a vision-language-action model built on Gemini 2.0 that adds physical actions as a new output modality for direct robot control (Parada, 2025). A companion model, Gemini Robotics-ER, handles embodied reasoning, and roboticists can connect it to their existing low-level controllers (Parada, 2025). DeepMind organized the release around generality, interactivity, and dexterity, and reported that the model replans when an object slips from its grasp or someone moves it (Parada, 2025).

Physical Intelligence's π_0 paper shows the research pattern investors are underwriting. The authors build a flow-matching architecture on a pretrained vision-language model so the robot inherits internet-scale semantic knowledge (Black et al., 2024). They train it across single-arm, dual-arm, and mobile manipulation platforms, then evaluate zero-shot performance and fine-tuning on tasks such as laundry folding, table cleaning, and box assembly (Black et al., 2024).

Hsu places π_0 , Gemini Robotics, and GR00T N1 inside the same primitive (Hsu, 2026b). In that primitive, large-scale pretraining amortizes the cost of seeing and understanding, and new effort goes into action.

The sequence continues across late 2025 and 2026. Physical Intelligence published $\pi^*0.6$ in November 2025, adding reinforcement learning from autonomous robot experience and expert corrections to a pretrained vision-language-action model (Physical Intelligence et al., 2025). The paper reports that this training more than doubles throughput on some of the hardest tasks and roughly halves the failure rate (Physical Intelligence et al., 2025). Those results come from the company's own task set and training pipeline. NVIDIA released GR00T N1.7 in early access with commercial licensing in March 2026 and previewed GR00T N2, with availability slated for the end of the year (NVIDIA, 2026). NVIDIA ties its strongest claim to the N2 world action model line rather than to N1.7 (NVIDIA, 2026). That claim holds that robots succeed at new tasks in new environments more than twice as often as with leading vision-language-action models. In July 2026, DeepMind introduced Gemini Robotics 2 and extended control from the upper body to whole humanoid bodies (Parada, 2026). Its own reporting shows medium to high success on whole-body and gripper tasks, while multi-finger dexterity remains challenging (Parada, 2026).

None of these systems retires the reliability problem. Hsu's arithmetic shows why, since a 95 percent success rate per step yields only about 60 percent across a 10-step chain when those step probabilities compound (Hsu, 2026b). Production environments demand far better than that (Hsu, 2026b). The calculation is a warning rather than a universal reliability model, because real robot failures can be correlated, but the operating lesson survives. Physical AI fails in centimeters, newtons, cycle time, damaged product, and human exposure rather than in tone.

The fact is that NVIDIA, Google DeepMind, and Physical Intelligence are productizing vision-language-action stacks, synthetic data pipelines, and post-training from experience (NVIDIA, 2025, 2026; Parada, 2025, 2026; Black et al., 2024; Physical Intelligence et al., 2025). Their strongest performance claims remain condition-specific and largely vendor-reported. The tactic is to require a written plan for any robotics or physical automation pilot before hardware procurement locks. That plan covers real teleoperation data, synthetic augmentation, post-training, baseline measurement, and the owner of safety validation. The measure is in-spec successful completions per operating hour, mean human interventions per operating hour, and total cost per successful completion against a classical automation baseline, all registered before the pilot clock starts.

What does Tesla's Optimus and Robotaxi record show about the deployment gap?

Tesla is placing one physical AI bet across two embodiments, a car and a humanoid, and its own filings now describe the company in those terms. Its fourth-quarter 2025 update called 2025 a year in Tesla's "transition from a hardware-centric business to a physical AI company" (Tesla, 2026c). Its 2025 preliminary proxy statement says Tesla "is leveraging FSD's core AI and vision systems to allow its Bots to perceive their environment and perform tasks" (Tesla, 2025b). At CVPR in June 2026, Tesla described "constructing foundation models for robotics" built as "large-scale multimodal models that control these robots in an end-to-end pixels-to-actuation fashion" (Tesla, 2026e). That places Tesla inside the same primitive as GR00T N1, Gemini Robotics, and π_0 , with a different data bet. Tesla describes its driving system as an "end-to-end foundation model trained on both customer and Robotaxi real-world data" (Tesla, 2026c). Tesla AI chief Ashok Elluswamy has also been reported describing a single world simulator for both the cars and Optimus (Humanoids Daily, 2025). The sources cited here describe that approach through investor letters, event pages, and press reports rather than through a technical paper with a stated evaluation.

The Optimus record shows how far a demonstration can sit from production work. Tesla's second-quarter 2024 update said Optimus "began performing tasks autonomously in one of our facilities," with battery handling as its first task (Tesla, 2024). That October, at the "We, Robot" event, Optimus units walked among guests, poured drinks, and held conversations. Reporting that followed found humans remotely operating much of that interaction, while the walking itself was described as AI (Bellan, 2024; Orland, 2024). One drink-serving unit told a guest, "Today, I'm assisted by a human. I'm not yet fully autonomous" (Orland, 2024). When Tesla showed an upgraded hand with 22 degrees of freedom that November, it confirmed that the demonstration was teleoperated (Lambert, 2024).

The volume record tells the same story in numbers. In January 2025, Musk said Tesla would probably not build 10,000 Optimus robots that year but would make several thousand, adding, "I'm confident they will do useful things" (Lambert, 2025). On the January 28, 2026 earnings call, he said Optimus remained in an "R&D phase," was not contributing to Tesla's manufacturing work in any "material way," and was in the factory for training (Lee, 2026). Tesla's update that day planned a Gen 3 unveiling in the first quarter, called it "our first design meant for mass production," and set an "eventual planned capacity of 1 million robots per year" (Tesla, 2026c). The unveiling had still not happened by late September (Lambert,

2026b). In July 2026, Musk warned that "Optimus production will be extremely slow at first, as everything is new," adding, "This is not like making a car" (Lambert, 2026a).

Tesla's July update said it had decommissioned the Model S and Model X lines at Fremont and was installing first-generation Optimus lines, with initial builds going to an "Optimus Academy for training data collection and further functionality development" (Tesla, 2026b). In September 2026, Electrek, citing The Information, reported output of several hundred units a week in August (Lambert, 2026b). The same report described hand and forearm assemblies still built by hand, and people familiar with the system said its AI "can't yet reliably handle a wide range of tasks" (Lambert, 2026b). The forearm is where Figure located its top hardware failure point at BMW (Figure AI, 2025). Tesla has not confirmed the weekly rate, so it stands as reported rather than disclosed.

The Robotaxi record shows why the word autonomous needs an operating mode attached to it. Tesla opened service in Austin on June 22, 2025 with an employee in the front passenger seat as a "safety monitor" (O'Kane & Korosec, 2025). Tesla says it began testing driverless vehicles in Austin in December 2025 and began removing the safety monitor from customer rides in January 2026 "on a limited basis" (Tesla, 2026c). Electrek then reported video evidence it read as trailing cars carrying safety monitors behind vehicles described as unsupervised (Lambert, 2026c). Tesla's March 2026 letter to Senator Markey gives the company's own account of the human layer. It states that the automated driving system is designed to perform the driving task "without the need for remote assistance" (Tesla, 2026d). Remote assistance operators may still "change the vehicle's trajectory or driving path," issue discrete commands, and, "as the final escalation maneuver," take temporary direct control at 2 mph or less, with an enforced maximum of 10 mph when the driving system grants direct access (Tesla, 2026d). The same letter describes "a fully autonomous rideshare service in Austin, Texas, since June 2025" (Tesla, 2026d), a description the launch record with its passenger-seat monitor does not match. Writing on launch day, Phil Koopman suggested that a more accurate description was probably "FSD(remote supervised)" (Koopman, 2025). No one in the driver's seat, no one in the car, no one following, and no one able to steer from a remote desk are four different claims.

By July 2026, Tesla listed unsupervised operation ramping in Austin, Dallas, Houston, Miami, Orlando, and Tampa, while its San Francisco Bay Area service ran with a safety driver (Tesla, 2026b). California's DMV listed Tesla Robotaxi LLC among holders of an autonomous vehicle testing permit with a driver as of

September 17, 2026 (California Department of Motor Vehicles, 2026a). On September 3, 2026, Tesla began commercial deployment of a small number of Cybercabs in Austin (National Highway Traffic Safety Administration, 2026b). NHTSA describes the Cybercab as "a vehicle lacking traditional human controls" (National Highway Traffic Safety Administration, 2026a). The agency opened an audit of the basis for Tesla's self-certification that the Cybercab meets federal motor vehicle safety standards. That audit includes "whether Tesla's compliance framework relied on determinations that certain standard FMVSS requirements are inapplicable to its automated vehicles" (National Highway Traffic Safety Administration, 2026a).

The vision stack Tesla says its robots share is under federal review as well. In March 2026, NHTSA upgraded its reduced-visibility investigation of FSD, which the agency describes as an advanced driver assistance system, to an engineering analysis covering an estimated 3,203,754 vehicles (National Highway Traffic Safety Administration, 2026c). The agency described FSD as relying "exclusively on vision-based cameras" and reported that Tesla's own post-incident analysis found the update to its degradation detection system, had it been installed at the time, "may have affected 3 of the 9 incidents" (National Highway Traffic Safety Administration, 2026c). NHTSA added that it did not know when the update was deployed or which vehicles carry it (National Highway Traffic Safety Administration, 2026c). It also recorded that Tesla described "internal data and labeling limitations," which the agency believes could have led to under-reporting of crashes (National Highway Traffic Safety Administration, 2026c).

The names attached to capability became a regulatory matter too. California's DMV found in December 2025 that Tesla's use of "autopilot" and "Full Self-Driving Capability" was misleading, and by February 2026 Tesla had stopped using "Autopilot" in its California marketing (California Department of Motor Vehicles, 2025, 2026b). Liability reached a jury as well. In February 2026, a federal court denied Tesla's motion for judgment as a matter of law or a new trial in a case arising from an April 2019 crash of a Model S equipped with Autopilot. That ruling left a \$242.57 million judgment in place, with Tesla assigned 33 percent of fault (*Benavides v. Tesla, Inc.*, 2026).

The capital record shows the bet is formal. Tesla shareholders approved the 2025 CEO Performance Award at the November 6, 2025 annual meeting (Tesla, 2025a), and its product goals include "1 Million Bots Delivered" and "1 Million Robotaxis in Commercial Operation" (Tesla, 2025b). The same proxy records Musk telling employees in March 2025 that Optimus could become "the biggest product of all

time by far" (Tesla, 2025b). Tesla's second-quarter 2026 capital expenditures reached \$5.8 billion, up 142 percent year over year, and free cash flow turned negative at \$1.1 billion (Tesla, 2026b). On September 29, 2026, Tesla entered credit agreements totaling \$30 billion, with no loans outstanding and proceeds available for general corporate purposes rather than earmarked for robots (Tesla, 2026a). Executive compensation, capital spending, and balance-sheet capacity now point toward physical AI while Tesla's own updates still describe Optimus output as training builds.

Independent readers divide sharply on the same record. After the 2024 event, Morgan Stanley's Adam Jonas wrote that the robots "relied on tele-ops (human intervention)," making the evening "more a demonstration of degrees of freedom and agility," while Canaccord Genuity's George Gianarikas answered, "So What!" (Bellan, 2024; Orland, 2024). In October 2025, Jonas wrote, "I'm callin' it. Autonomous cars are solved" (Karaahmetovic, 2025). A month later, Missy Cummings, a George Mason University professor who directs the Mason Autonomy and Robotics Center, told EE Times that "vision-only self-driving cars are never, ever, ever going to happen" (Pelé, 2025). By August 2026, Andrew Percoco, who took over Morgan Stanley's Tesla coverage from Jonas, had reportedly set an evidence list for Optimus (EVSHIFT, 2026). The list asks for a production-ready robot shown publicly rather than under controlled conditions, factory operation without constant human supervision, credible manufacturing cost, outside customer orders, and a realistic path to mass production (EVSHIFT, 2026). That list reads like a deployment ledger, and it applies the same test this article applies to every physical AI claim.

The fact is that Tesla's filings, its letter to a United States senator, and its regulators' records show a physical AI program whose claims move faster than its disclosed operating evidence (Tesla, 2026c, 2026d; National Highway Traffic Safety Administration, 2026a, 2026c). Human assistance appears at the demonstration, in the passenger seat, at the remote desk, and, by one Electrek report, in a trailing car (Orland, 2024; O'Kane & Korosec, 2025; Tesla, 2026d; Lambert, 2026c). The tactic is to log every physical AI claim against its operating mode, whether onboard autonomy, an in-vehicle safety operator, chase or remote monitoring, a remote path change, remote direct control, or teleoperated task execution, each with a date and a source. The measure is the share of autonomy claims in executive reviews that carry a dated operating mode and a source class, with a target of 100 percent.

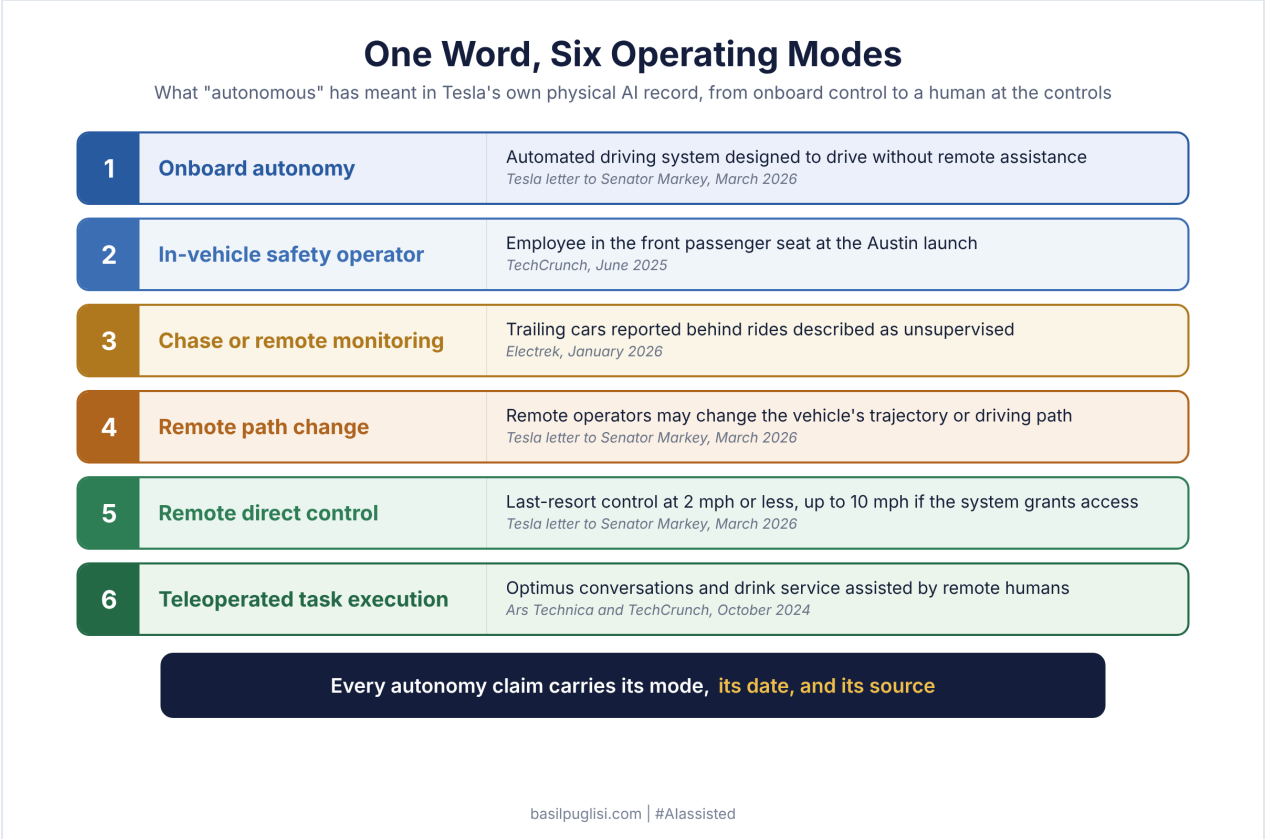


Figure 1. Six operating modes that can sit behind the word autonomous in Tesla's public record, each with its date and source.

Where should strategy teams place capital before the humanoid demos distract them?

Waiting for a general-purpose humanoid to clear every hallway is the wrong strategy. The stronger move is to buy closed-loop advantage where sensing, action, and economics already meet. NVIDIA positions GR00T N1 for material handling, packaging, and inspection, the repetitive manipulation categories where foundation-model robotics is already being trained and measured (NVIDIA, 2025). Autonomous science deserves a parallel lane because a self-driving laboratory turns physical experiments into structured, causal, empirically verified training signal that internet text cannot supply (Hsu, 2026b). Interface investments matter for a related reason (Hsu, 2026b). AR glasses, voice-first wearables, and, over a longer horizon, neural interfaces can instrument human physical experience and expand the sensory surface available to AI systems.

The deployment gap changes the burden of proof, because classical automation starts as the incumbent rather than as the technology that must defend itself. Suppose a conventional integration without foundation-model post-training matches the learning-based pilot on throughput, quality, intervention rate, and validated safety at lower total cost over a full operating quarter. In that case, the foundation-model premium is not justified for that cell. The learning-based system earns the spend only when the ledger beats the alternative.

The hardware ledger belongs in the same decision. Figure's own report on its Figure 02 deployment at BMW says the robots ran more than 1,250 hours and loaded more than 90,000 parts (Figure AI, 2025). Across the 11-month deployment, they contributed to the production of more than 30,000 vehicles (Figure AI, 2025). The same report names the forearm as the robot's top hardware failure point at BMW and describes a redesign of the wrist electronics for Figure 03 (Figure AI, 2025). Even a favorable deployment report therefore shows why physical AI cannot be evaluated as software alone. Models can be updated without wearing out, while joints, cables, sensors, thermal systems, and power electronics cannot.

That comparator also exposes a resilience requirement. A learned system can outperform the incumbent and still leave the operation more fragile. That happens when the cell cannot continue after the model, network connection, provider, or post-training pipeline becomes unavailable. The Council for Humanity proposal addresses exactly this condition (Puglisi, 2026a). It calls for AI-integrated critical infrastructure to maintain and regularly test AI-independent operational capability, so an AI interruption degrades performance rather than collapsing the system, which the author calls the Digital Resilience Requirement (Puglisi, 2026a). Applied at the cell level, the same principle asks what still works when the learned layer is unavailable. The fallback may be a deterministic controller, a classical automation mode, a manual procedure, or another safe degraded state. The goal is a tested answer to that question, and it carries no obligation to preserve yesterday's process forever.

Factics scores a physical AI opportunity the way it scores any intervention. The first question asks what is known about the current failure or cost, and the second asks what exact intervention the system will perform. The third asks which measure, defined before the work begins, will show that the intervention worked (Puglisi, 2026c). A pilot whose only measure is a model release or a demo view is funding a narrative. A pilot measured in in-spec units per hour, defined quality, intervention rate, safety events, and full cost is funding operations.

Talent follows the same map, and the mechanical system stays on it. The scarce profile connects model post-training to embodiment constraints, simulation coverage, safety controllers, integration engineering, and plant reality. Hsu argues that value in robotics increasingly moves toward models, training infrastructure, and data flywheels as learned policies become standard (Hsu, 2026b). Figure's BMW report supplies the counterweight, since hardware reliability still changed the design (Figure AI, 2025). Organizations that staff physical automation as pure mechanical integration risk renting intelligence from someone else. Organizations that treat mechanics as a solved commodity risk discovering failure at the wrist, cable, sensor, or safety boundary.

The deployment gap also creates predictable governance pressure. *Governing AI: When Capability Exceeds Control* describes the systematic tendency for profit maximization, competitive pressure, and shareholder returns to prioritize capability advancement over safety validation, creating predictable governance failures absent mandatory accountability structures, which the author calls the Economic Override Pattern (Puglisi, 2025b). Physical AI makes that mechanism tangible. A strong demonstration creates pressure to buy, and the labor story creates pressure to scale. Sunk hardware and integration costs then create pressure to continue even when the local evidence weakens. Preregistered kill rules matter because they write the stop condition into the investment decision before those incentives compound.

The labor question follows the same architecture. The author's earlier work distinguishes a replacement architecture, Culture 1, from an augmentation architecture, Culture 2, even when both run on the same frontier models (Puglisi, 2026b). In physical work, Culture 1 removes the human layer, while Culture 2 moves dangerous, repetitive, strength-intensive, and precision-sensitive work to the machine and keeps human judgment where consequence requires it. Physical AI makes that choice concrete, because a robot can reduce human exposure without making labor displacement the measure of success. The operating question becomes what work changed, what human capability increased, and whether the combined system became safer and more productive.

The fact is that near-term value concentrates where closed-loop tasks, combined synthetic and real data recipes, clear operational measures, and an incumbent baseline already exist (Hsu, 2026a, 2026b; Parada, 2026; Legate-Yang & Massenkoff, 2026). Humanoid generality remains an unfinished race on reliability and economics, and deployment incentives can push organizations across that gap too early (Puglisi, 2025b). The tactic is to fund only pilots whose evaluation team

can preregister, baseline, and audit the work. Each operational cell pairs with the data or simulation work meant to improve it, keeps a tested fallback mode, and states whether it is designed primarily for replacement or augmentation. The measure is payback against the priced classical alternative, combined with staged gates on intervention, quality, and fallback recovery. Each gate closes on accept, modify, or reject as the evidence accumulates.

What weekly cadence keeps physical AI from becoming theater?

On Monday, the team reviews the pilot scoreboard and refuses vanity metrics, counting completed tasks, failure modes, human interventions, quality escapes, and safety events. On Tuesday, it briefs the next batch of data collection or synthetic generation against those failure modes rather than against the marketing story. On Wednesday, the learning team decides whether retraining or post-training is warranted, working from a written hypothesis about which distribution gap the next batch should close. A weekly governance cadence does not mean a new model enters production every week. Any change that requires engineering validation or a new safety review stays out of the operating cell until that work is complete.

On Thursday, the team evaluates the candidate system on held-out objects, lighting, layouts, and other preregistered stress conditions. Generality claims break in corner cases, and DeepMind's 2025 release treats real-world surprises as a normal condition of operation (Parada, 2025). On Friday, a named person reads the preregistered measures and decides to accept and scale, modify and redesign, or reject and stop, and that person answers for the decision.

Checkpoint-Based Governance treats that moment as a human checkpoint (Puglisi, 2025a, 2026d). The outcome is accept, modify, or reject, and authority sits with a named human rather than the pipeline (Puglisi, 2026d). The checkpoint does not replace a safety-rated controller, a risk assessment, or an engineering signoff. It governs the decision over whether the technical evidence, safety case, and economics justify what happens next.

Operators already run parts of this rhythm. Tesla told Senator Edward Markey that all remote assistance actions on its robotaxis are logged and that the company "conducts weekly performance audits for selected sessions to assess remote recovery metrics" (Tesla, 2026d). An audit records what happened, while a checkpoint decides what happens next, and a physical AI cadence needs both.

Physical AI makes one governance question harder rather than less relevant. The question asks whether a named human holds binding authority to accept, modify, or reject an output and remains accountable for the result, which the author calls the Named-Human Test (Puglisi, 2026e). A physical agent leaves that question in place and raises the consequence, because physical AI is agentic AI with kinetic consequences. The checkpoint may sit upstream of millisecond control, as the test's high-velocity provision allows (Puglisi, 2026e). A qualified human still authorizes the policy, defines the operating constraints, accepts the residual risk, and can reject the deployment at any checkpoint.

The fact is that physical AI changes faster than industrial validation can safely absorb every model update, while the decision to release more capital remains a governance act. The tactic is to run a weekly evidence cadence without assuming a weekly deployment cadence. The measure is the percentage of model or policy changes that enter production only after their preregistered evaluation and required safety review are complete, with a target of 100 percent.

How should leaders read physical AI claims without overclaiming them?

NVIDIA's labor-shortage framing, DeepMind's dexterity videos, Physical Intelligence's throughput results, and investor essays about frontier systems all count as directional evidence about a stack in motion (NVIDIA, 2025; Parada, 2025, 2026; Physical Intelligence et al., 2025; Hsu, 2026b). None of it warrants the conclusion that a given facility gets a generalist robot next spring. The GR00T N1 gain is tied to a specific training mix and NVIDIA's own evaluation (Vadrevu & Omotuyi, 2025). Gemini Robotics generalization claims rest on DeepMind's reported evaluations and partner settings (Parada, 2025). The π_0 and $\pi^*0.6$ results are research outcomes under stated data and embodiment conditions (Black et al., 2024; Physical Intelligence et al., 2025). Each source carries its author's incentives and test conditions. Platform evidence informs the bet, and the local preregistered measure decides the money.

Three mechanisms are worth borrowing directly. The first treats vision-language-action models as a software layer that needs a data flywheel, which turns the purchase into an ongoing program rather than a one-time robot buy (Hsu, 2026b; NVIDIA, 2025). The second budgets synthetic and simulated experience as first-class training infrastructure, because real robot trajectories are expensive to collect and do not scale the way internet text did (Hsu, 2026b; Vadrevu & Omotuyi, 2025). The third separates semantic reasoning from engineered safety. DeepMind

describes embodied reasoning connected to low-level, safety-critical controllers specific to each robot (Parada, 2025), while industrial deployments remain subject to safety requirements that sit outside any model benchmark.

That distinction matters. ISO 10218-1:2025 covers safety requirements for industrial robots, and ISO 10218-2:2025 covers industrial robot applications and robot cells, including integration, commissioning, operation, and maintenance (International Organization for Standardization, 2025a, 2025b). ISO/TS 15066:2016 remains current for collaborative industrial robot systems while a successor is under development (International Organization for Standardization, 2016). A vision-language-action model that refuses an unsafe instruction can be useful. That behavior still cannot substitute for the engineered safety functions, integration controls, and risk assessment required around an industrial robot.

DeepMind's 2026 release adds a separate signal worth tracking. Its ASIMOV-Agentic benchmark measures whether the reasoning agent refuses unsafe tool calls from the action model and requests human intervention when it is uncertain (Parada, 2026). That is evidence about semantic safety behavior, not certification of the robot cell. A robot that asks for a person when it is unsure is compatible with a governed checkpoint, but the checkpoint sits above the safety system rather than replacing it.

The fact is that published physical AI progress is system-level, condition-specific, and split across three layers that should not be collapsed: model capability, deployment economics, and functional safety (Hsu, 2026a, 2026b; Parada, 2025, 2026; Legate-Yang & Massenkoff, 2026; International Organization for Standardization, 2025a, 2025b). The tactic is to attach a one-page Factics card to every executive demo review. The card states the source claim and classifies it as vendor, investor, operator, independent research, or standards evidence. It records the operating mode behind the claim, the local baseline, and the deployment gap, then names the kill rule and the decision owner. The measure is the percentage of triggered checkpoints that close on a documented accept, modify, or reject decision by the named human, audited quarterly.

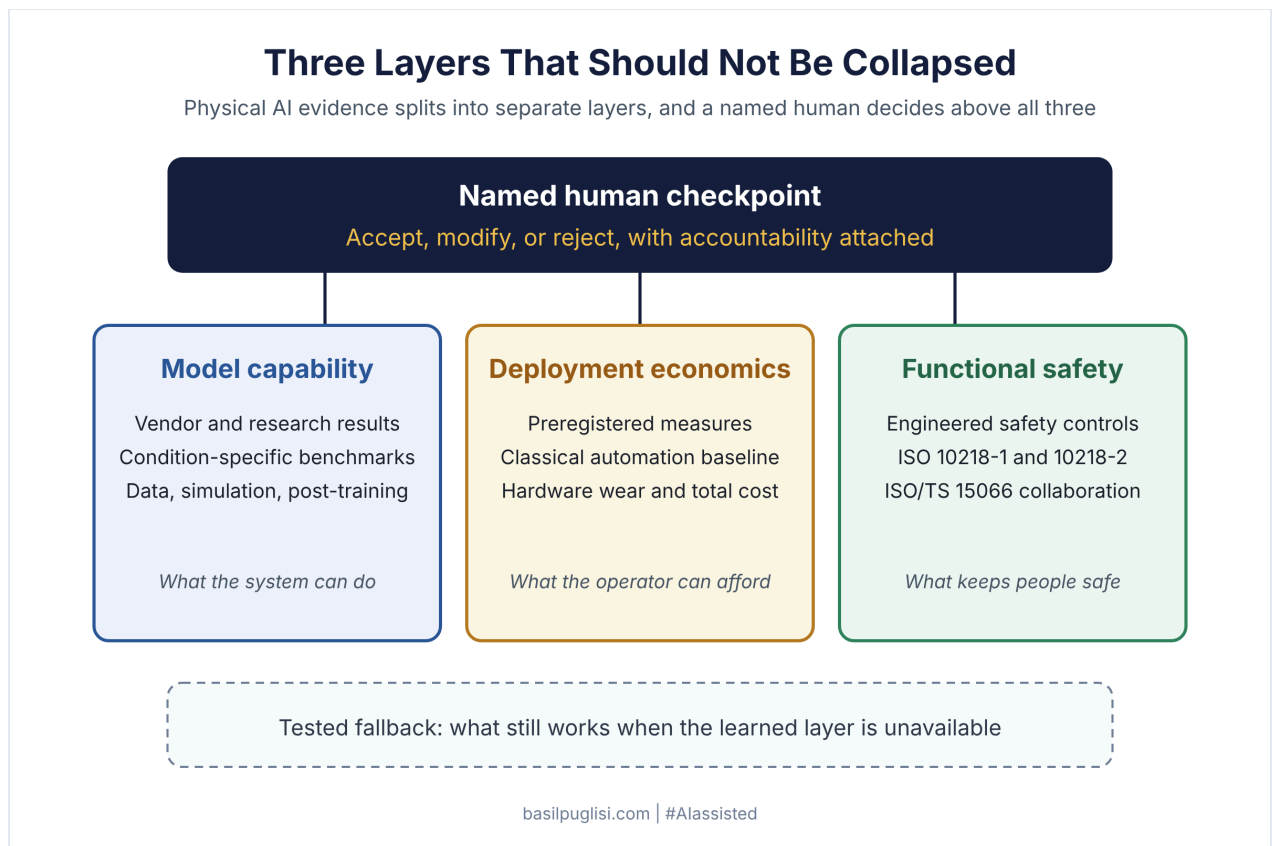


Figure 2. Model capability, deployment economics, and functional safety stay separate, and a named human decides above all three.

What should leaders watch next?

Reliability across multi-step tasks remains the deployment bottleneck Hsu flags, and reinforcement learning from robot experience is one route the field is testing to move it (Hsu, 2026b; Physical Intelligence et al., 2025). Open models, synthetic data blueprints, and shared physics engines keep lowering parts of the cost of post-training new embodiments. NVIDIA has placed GR00T N2 on that path, with availability slated for the end of 2026 (NVIDIA, 2026). Agent-level safety behavior now has a published benchmark behind it, while the industrial safety case still belongs to engineered controls and applicable standards (Parada, 2026; International Organization for Standardization, 2025a, 2025b).

Tesla's open items belong on the same watch list. NHTSA's Cybercab audit will show how far a vehicle without traditional human controls can rest on a manufacturer's own reading of which federal standards apply (National Highway Traffic Safety Administration, 2026a). The FSD engineering analysis bears directly on the camera-based stack Tesla says its robots draw on (National Highway Traffic Safety Administration, 2026c; Tesla, 2025b). For Optimus, the signal worth tracking

is the first published unit count, useful-work measure, or outside order, because a production line that feeds a training academy is still a data program rather than a labor product (Tesla, 2026b).

The economics deserve equal attention. Anthropic's estimate that robots are cost-competitive for only 0.3 percent of job tasks does not show that the platform shift fails (Legate-Yang & Massenkoff, 2026). It shows that the technical platform and the economic platform are moving at different speeds. That gap is where operators either create advantage or burn capital.

Competitors that build data loops, closed-loop evaluation, mechanical reliability, and safety evidence together will learn faster than organizations still staffing slide decks about copilots or buying robots because a demo went viral. For each pilot, the minimum record preserves the model and policy version, source provenance for load-bearing claims, data lineage, evaluation conditions, safety review, intervention history, decision owner, and final disposition. That record connects the demonstration to the audit trail. It gives a later reviewer enough evidence to reconstruct why the deployment was accepted, modified, or rejected, and it keeps the platform thesis answerable to the deployment ledger.

References

- Bellan, R. (2024, October 14). *Tesla Optimus bots were controlled by humans during the 'We, Robot' event*. TechCrunch. <https://techcrunch.com/2024/10/14/tesla-optimus-bots-were-controlled-by-humans-during-the-we-robot-event/>
- Benavides v. Tesla, Inc.*, No. 1:21-cv-21940 (S.D. Fla. Feb. 19, 2026) (order denying amended renewed motion for judgment as a matter of law or new trial). <https://law.justia.com/cases/federal/district-courts/florida/flsdce/1:2021cv21940/593426/612/>
- Black, K., Brown, N., Driess, D., Esmail, A., Equi, M., Finn, C., Fusai, N., Groom, L., Hausman, K., Ichter, B., Jakubczak, S., Jones, T., Ke, L., Levine, S., Li-Bell, A., Mothukuri, M., Nair, S., Pertsch, K., Shi, L. X., . . . Zhilinsky, U. (2024). *π0: A vision-language-action flow model for general robot control*. arXiv. <https://arxiv.org/abs/2410.24164>
- California Department of Motor Vehicles. (2025, December 16). *DMV finds Tesla violated California state law*. <https://www.dmv.ca.gov/portal/news-and-media/news-releases/dmv-finds-tesla-violated-california-state-law/>
- California Department of Motor Vehicles. (2026a). *Autonomous vehicle testing permit holders*. Retrieved October 7, 2026, from <https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-testing-permit-holders/>
- California Department of Motor Vehicles. (2026b, February 17). *Tesla takes corrective action to avoid DMV suspension*. <https://www.dmv.ca.gov/portal/news-and-media/tesla-takes-corrective-action-to-avoid-dmv-suspension/>

- EVSHIFT. (2026, August 16). *Morgan Stanley says Tesla has to prove the robots are real*. <https://www.evshift.com/510227/morgan-stanley-says-tesla-has-to-prove-the-robots-are-real/>
- Figure AI. (2025, November 19). *F.02 contributed to the production of 30,000 cars at BMW*. <https://www.figure.ai/news/production-at-bmw>
- Hsu, O. (2026a, January 13). *The physical AI deployment gap*. Andreessen Horowitz. <https://www.a16z.news/p/the-physical-ai-deployment-gap>
- Hsu, O. (2026b, April 15). *Frontier systems for the physical world*. Andreessen Horowitz. <https://a16z.com/frontier-systems-for-the-physical-world/>
- Humanoids Daily. (2025, November 3). *The physical AI bottleneck: Comparing the data strategies of 1X, Figure, Tesla, and Neura*. <https://www.humanoidsdaily.com/news/the-physical-ai-bottleneck-comparing-the-data-strategies-of-1x-figure-tesla-and-neura>
- International Organization for Standardization. (2016). *ISO/TS 15066:2016: Robots and robotic devices, collaborative robots*. <https://www.iso.org/standard/62996.html>
- International Organization for Standardization. (2025a). *ISO 10218-1:2025: Robotics, safety requirements, Part 1: Industrial robots*. <https://www.iso.org/standard/73933.html>
- International Organization for Standardization. (2025b). *ISO 10218-2:2025: Robotics, safety requirements, Part 2: Industrial robot applications and robot cells*. <https://www.iso.org/standard/73934.html>
- Karaahmetovic, V. (2025, October 27). *Morgan Stanley's Jonas declares robotaxi breakthrough: "I'm callin' it"*. Investing.com. <https://www.investing.com/news/stock-market-news/morgan-stanleys-jonas-declares-robotaxi-breakthrough-im-callin-it-4310049>
- Koopman, P. (2025, June 22). *Tesla robotaxis go live in Austin*. Autonomous System Safety. <https://philkoopman.substack.com/p/tesla-robotaxis-go-live-in-austin>
- Lambert, F. (2024, November 29). *Tesla unveils upgraded Optimus robot hand, but impressive demo is again teleoperated*. Electrek. <https://electrek.co/2024/11/29/tesla-unveils-upgraded-optimus-robot-hand-but-impressive-demo-is-again-teleoperated/>
- Lambert, F. (2025, January 31). *Elon Musk says Tesla aims to build 10,000 Optimus robots this year*. Electrek. <https://electrek.co/2025/01/31/elon-musk-says-tesla-aims-to-build-10000-optimus-robots-this-year/>
- Lambert, F. (2026a, July 2). *Elon Musk shuts down '4D chess' theory on Tesla Optimus production*. Electrek. <https://electrek.co/2026/07/02/musk-shuts-down-optimus-4d-chess-theory/>
- Lambert, F. (2026b, September 25). *Tesla ramps Optimus to hundreds a week, but the robots can't generalize*. Electrek. <https://electrek.co/2026/09/25/tesla-optimus-production-ramp-hands-ai-generalization-problems/>
- Lambert, F. (2026c, January 22). *Tesla starts Robotaxi rides without safety monitor in Austin: What you need to know*. Electrek. <https://electrek.co/2026/01/22/tesla-starts-robotaxi-rides-without-safety-monitor-in-austin-what-you-need-to-know/>
- Lee, L. (2026, January 29). *6 biggest takeaways from Tesla's Q4 earnings call*. Business Insider. <https://www.businessinsider.com/tesla-q4-earnings-call-summary-robotaxi-optimus-ai5-chips-2026-1>

- Legate-Yang, R., & Massenkoff, M. (2026, September 30). *What work can robots do?* Anthropic. <https://www.anthropic.com/research/what-work-can-robots-do>
- National Highway Traffic Safety Administration. (2026a, September 4). *NHTSA opens investigation into Tesla Cybercab self-certification following Austin deployment*. <https://www.nhtsa.gov/press-releases/investigation-tesla-cybercab-self-certification>
- National Highway Traffic Safety Administration. (2026b, September 3). *OVSC resume, Audit Query AQ26002: Tesla Cybercab FMVSS certification* [Hosted copy]. CBS Austin. <https://cbsaustin.com/resources/pdf/35e6ebe1-22af-4f48-836e-5bb77e815a17-OpenAuditQuery.pdf>
- National Highway Traffic Safety Administration. (2026c, March 18). *ODI resume, Engineering Analysis EA26002: FSD collisions in reduced roadway visibility conditions*. <https://static.nhtsa.gov/odi/inv/2026/INOA-EA26002-10023.pdf>
- NVIDIA. (2024, March 18). *NVIDIA announces Project GR00T foundation model for humanoid robots and major Isaac Robotics platform update*. NVIDIA Newsroom. <https://nvidianews.nvidia.com/news/foundation-model-isaac-robotics-platform>
- NVIDIA. (2025, March 18). *NVIDIA announces Isaac GR00T N1, the world's first open humanoid robot foundation model, and simulation frameworks to speed robot development*. NVIDIA Newsroom. <https://nvidianews.nvidia.com/news/nvidia-isaac-gr00t-n1-open-humanoid-robot-foundation-model-simulation-frameworks>
- NVIDIA. (2026, March 16). *NVIDIA and global robotics leaders take physical AI to the real world*. NVIDIA Newsroom. <https://nvidianews.nvidia.com/news/nvidia-and-global-robotics-leaders-take-physical-ai-to-the-real-world>
- O'Kane, S., & Korosec, K. (2025, June 22). *Tesla launches robotaxi rides in Austin with big promises and unanswered questions*. TechCrunch. <https://techcrunch.com/2025/06/22/tesla-launches-robotaxi-rides-in-austin-with-big-promises-and-unanswered-questions/>
- Orland, K. (2024, October 15). *Reports: Tesla's prototype Optimus robots were controlled by humans*. Ars Technica. <https://arstechnica.com/ai/2024/10/reports-teslas-prototype-optimus-robots-were-controlled-by-humans/>
- Parada, C. (2025, March 12). *Gemini Robotics brings AI into the physical world*. Google DeepMind. <https://deepmind.google/blog/gemini-robotics-brings-ai-into-the-physical-world/>
- Parada, C. (2026, July 30). *Gemini Robotics 2 brings whole body intelligence to robots*. Google DeepMind. <https://deepmind.google/blog/gemini-robotics-2-brings-whole-body-intelligence-to-robots/>
- Pelé, A.-F. (2025, November 21). *Every self-driving car company needs human babysitters*. EE Times. <https://www.eetimes.com/every-self-driving-car-company-needs-human-babysitters/>
- Physical Intelligence, Amin, A., Aniceto, R., Balakrishna, A., Black, K., Conley, K., Connors, G., Darpinian, J., Dhabalia, K., DiCarlo, J., Driess, D., Equi, M., Esmail, A., Fang, Y., Finn, C., Glossop, C., Godden, T., Goryachev, I., Groom, L., . . . Zhou, Z. (2025). *$\pi^*0.6$: A VLA that learns from experience*. arXiv. <https://arxiv.org/abs/2511.14759>
- Puglisi, B. C. (2012, November 27). *Digital Facts: Twitter*. Digital Media Press. <https://www.magcloud.com/browse/issue/471388>
- Latest update: Puglisi, B. C. (2026c, September 2). *Facts: The method behind valuable content and trusted AI*. basilpuglisi.com. <https://basilpuglisi.com/facts/>

- Puglisi, B. C. (2025a, September 23). *Checkpoint-Based Governance: An implementation framework for accountable human-AI collaboration*. basilpuglisi.com. <https://basilpuglisi.com/checkpoint-based-governance-an-implementation-framework-for-accountable-human-ai-collaboration-v2-drafting/>
- Latest update: Puglisi, B. C. (2026d, September 11). *What is Checkpoint-Based Governance? Human authority*. basilpuglisi.com. <https://basilpuglisi.com/cbg/>
- Puglisi, B. C. (2025b). *Governing AI: When capability exceeds control*. ISBN 9798349677687. <https://basilpuglisi.com/governing-ai-when-capability-exceeds-control/>
- Puglisi, B. C. (2026a, February 2). *Council for Humanity*. basilpuglisi.com. <https://basilpuglisi.com/council-for-humanity/>
- Puglisi, B. C. (2026b, May 31). *Did AI write Magnifica Humanitas? Pope Leo XIV was the author, but what was the governance method?* basilpuglisi.com. <https://basilpuglisi.com/did-ai-write-magnifica-humanitas/>
- Puglisi, B. C. (2026e, June 4). *Why agentic AI was always going to fail*. basilpuglisi.com. <https://basilpuglisi.com/why-agentic-ai-was-always-going-to-fail/>
- Tesla, Inc. (2024, July 23). *Q2 2024 update*. https://ir.tesla.com/_flysystem/s3/sec/000162828024032603/tsla-20240723-gen.pdf
- Tesla, Inc. (2025a, November 7). *Current report on Form 8-K (date of report November 6, 2025): Submission of matters to a vote of security holders*. U.S. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/1318605/000110465925108507/tm2530590d1_8k.htm
- Tesla, Inc. (2025b, September 5). *Preliminary proxy statement (Schedule 14A)*. https://ir.tesla.com/_flysystem/s3/sec/000110465925087598/tm252289-4_pre14a-gen.pdf
- Tesla, Inc. (2026a, September 29). *Current report on Form 8-K: Entry into a material definitive agreement*. U.S. Securities and Exchange Commission. <https://www.sec.gov/Archives/edgar/data/1318605/000162828026063820/tsla-20260929.htm>
- Tesla, Inc. (2026b, July 22). *Q2 2026 update*. https://ir.tesla.com/_flysystem/s3/sec/000162828026049213/tsla-20260722-gen.pdf
- Tesla, Inc. (2026c, January 28). *Q4 and FY 2025 update*. https://ir.tesla.com/_flysystem/s3/sec/000162828026003837/tsla-20260128-gen.pdf
- Tesla, Inc. (2026d, March 26). *Letter to Senator Edward J. Markey on remote assistance for autonomous vehicles*. In *Company responses to remote assistance operator letter*. Office of Senator Edward J. Markey. <https://www.markey.senate.gov/download/company-responses-to-rao-letter?download=1>
- Tesla, Inc. (2026e, June). *Tesla AI at CVPR 2026*. <https://www.tesla.com/event/tesla-x-cvpr-2026>
- Vadrevu, K. M., & Omotuyi, O. (2025, March 18). *Accelerate generalist humanoid robot development with NVIDIA Isaac GR00T N1*. NVIDIA Technical Blog. <https://developer.nvidia.com/blog/accelerate-generalist-humanoid-robot-development-with-nvidia-isaac-gr00t-n1/>

Disclaimer

I am not a lawyer, and this article does not provide legal advice. This is thought research and governance analysis based on public sources, cited materials, and human-AI review. It is intended to help executives, practitioners, insurers, and governance teams think more clearly about AI risk, liability exposure, and documentation practices. Readers should not rely on this article as a legal opinion, compliance determination, or substitute for qualified counsel. Any organization facing a legal, regulatory, contractual, or insurance question should consult its own attorney, broker, or professional adviser before acting.

The Other AI: Audio Briefings on Augmented Intelligence and AI Governance

[Spotify](#) | [Apple Podcasts](#) | [Amazon Music](#) | [YouTube Playlist](#)

Basil C. Puglisi, MPA

A Human-AI Collaboration

#AIassisted using the HAIA Ecosystem | CC BY-NC-SA 4.0

Free for personal, educational, and noncommercial research use with attribution. Commercial exploitation, paid productization, and enterprise commercialization require separate permission and licensing.