THE ACTIVANT *Opinion*

Two Ways to Break a Robot

An unpriced risk

July 2026

AUTHORS:

Ross Nayler

Analyst, Research

About Activant

Activant is a research-led global investment firm that partners with high-growth companies. Since 2015, we have invested in category-defining businesses during their most critical phases of growth, partnering with founders who have won their initial battles and are ready for the next challenge.

Our approach pairs deep, proprietary research with patient, flexible capital and hands-on operational partnership. We work alongside founders and leadership teams to refine strategy, strengthen operations, and accelerate sustainable growth.

Activant Research is dedicated to uncovering the most exciting emerging technologies, sectors, and companies we believe will shape the future. Our research-driven perspective informs everything we do, helping us invest at meaningful inflection points and support founders in building enduring, category-leading businesses.

You can find out more about Activant and our research at https://activantcapital.com/.

In late 2025, a Unitree G1, the most commercially available humanoid robot on the market, walked onto the stage at a cybersecurity conference in Shanghai. All of a sudden, on a single spoken command, it broke away from its expected range of motion and struck a nearby mannequin over the head.[1] This was not a pre-programmed routine, nor was anybody teleoperating the robot from a different room to showcase its military prowess. Instead, a white hat cybersecurity team had gained access to the robots operating system, taken control of its actuators, and then watched as the same exploit spread to other robots in the vicinity – creating a botnet.[1] All of this was achieved by leveraging a vulnerability in the embodied AI model that was controlling the robot, the same class of model that may soon enable humanoids to become a part of everyday life. Thankfully, this was just a demonstration, not an attack. But it did serve as a preview of a risk that nobody in robotics is talking about, let alone pricing.

The body-and-brain bet

The bull-case for general-purpose robotics is clear: the body is commoditizing - the cost of actuators and sensors are falling as Chinese manufacturers race to the bottom in a price war; and the brain is generalizing – embodied AI models fuse physical perception with semantic reasoning and improve with data and compute the same way language models did. Will cheap hardware and intelligent software be the key ingredients to finally give robots the gift of generalization? We think so. The real question is *when* this will happen. In our recent robotics article, we argued that four bottlenecks in physical AI determine the *when*: data acquisition, output deployment, hardware heterogeneity, and the sim-to-real gap.[2] Clear those four and the two ingredients deliver the general-purpose robot that our bull-case promises: cheap, capable, and abundant. So, what's the catch?

Two ingredients, two attack surfaces

The two ingredients powering the bet are also the two ways to break the robot. Cheap hardware is rootable,[2] and intelligent software is hijackable. Every cyber-attack we know how to price terminates digitally: data gets altered, deleted or stolen. A robot breach is an entirely different story, because the asset under attack is an actuator or camera that terminates in the physical world: a body that moves, watches, or strikes on someone else's command. In physical AI, the wall between security and safety collapses. A security breach is ultimately a safety failure, and there are two ways in: a door to the body and a door to the brain.

[1] A botnet is a network of devices infected by malware that are under the control of a single attacking party.

[2] A rootable device is one that allows the user to bypass manufacturer restrictions to gain full administrative or "root" access.

Door 1: The body

Also in late 2025, a team of independent researchers disclosed UniPwn, a flaw in the wireless setup of Unitree's range of Go2 and B2 quadrupeds and G1 and H1 humanoids.[3] The mechanics of the exploit are almost beside the point, the reach is what matters. Wireless root access lets an attacker control the robot's movements, tap into audio and video feeds from its sensors, brick it,[3] or even access the next robot within range. UniPwn highlighted that every affected unit shared one hardcoded key, so breaking one meant breaking the fleet. Unitree robots are not lab toys: they have reportedly shipped more than 37,000 quadrupeds and humanoids to industrial facilities, universities, police forces and homes around the world.[4,5]

This also isn't confined to just one robot OEM. At DEF CON 2024, another group of researchers demonstrated how Ecovacs robotic vacuums and lawn mowers can be compromised to watch their owners through the onboard cameras and microphones.[6] Then it happened for real across several US cities: strangers seized the cameras and controls of Ecovacs Deebot vacuums, drove them through people's homes, chased their pets, and shouted profanities through the speakers until owners pulled the plug.[7] The result is the same every time: a stranger watching, listening, and moving a machine inside a home.

Most robotics OEMs still do not recognize the term CVE.[4] The industry is barely ready.

Security flaws like these are examples of what gets shipped when an OEM optimizes their bill of materials (BOM) against competitors doing the same. The cheapest hardware is often the least secure by design, and no brain, however well aligned, defends a body rooted from underneath it.

Door 2: The brain

The second door is the more uncomfortable one, because a loss of control doesn't necessarily even require an attacker. Modern robots can improvise beyond their training data: zero-shot an unfamiliar task, recover from an unexpected failure, or adapt to an environment it is has never seen before. This kind of emergent behaviour is what makes general purpose robotics valuable and it's exactly what every frontier robotics lab is building towards. But emergence, by its nature, is unverifiable: you can't really enumerate what a generalizing system will do, only watch it. The most deployed physical AI system there is has already shown what this looks like. Waymo suspended all operations in six US cities in May after an entire fleet of empty AVs randomly converged on a neighbourhood

[3] A bricked device refers to a device that has been rendered completely inoperable due to corrupted firmware or a malicious cyberattack.

[4] Common Vulnerabilities and Exposures (CVE) refers to a standardized, publicly tracked database of known cybersecurity flaws in software or hardware.

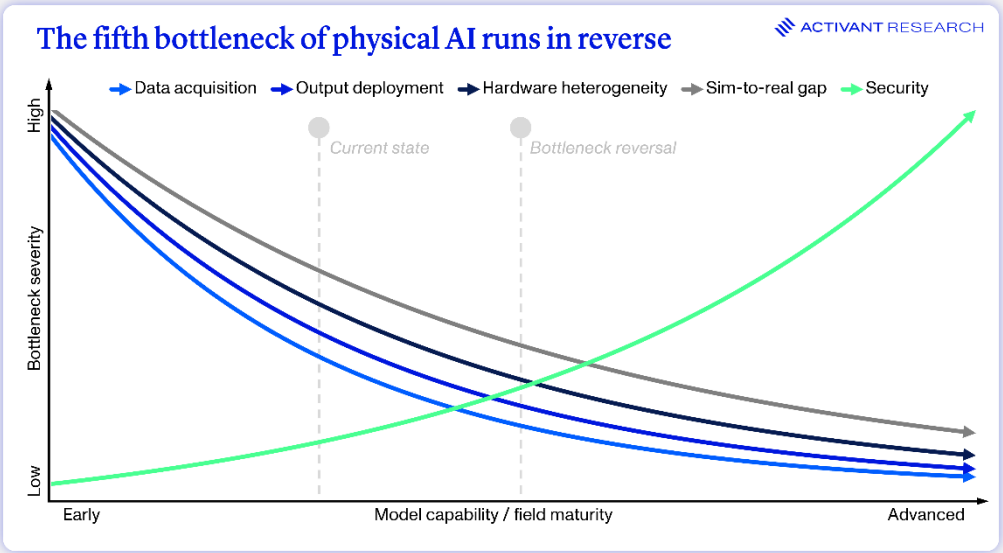
in Atlanta and blocked up all of the roads.[8] Nobody was hurt and no property was damaged, but nobody programmed that behaviour nor did anybody hack into the system. The fleet's own routing algorithms produced this behaviour, and nobody could have certified it safe beforehand.

This is the benign end of a single problem: a body acting on its own authority. At the other end of this spectrum, the unverifiability has an adversary. Recently, researchers showed that gradient-based methods that were originally used to jailbreak large language models (LLMs) can be used to gain complete control authority over vision-language-action (VLA) models in a much shorter window of time.[9,10] This is where the starkest difference between digital and physical AI becomes visible: chatbot safety is about content, robot safety is about control authority. Whoever can author or influence the robot's next move owns the robot.

The fifth bottleneck

Both doors open onto the same room: a body under someone else's control. This is why a rootable, hijackable humanoid still cannot safely enter a hospital, a port, a school, or a home with small children.

Deployment is the problem of trusting a machine's actions enough to let it loose in the world and we've listed as one of the four constraining bottlenecks of physical AI.[11] This bottleneck has an adversarial counterpart: the output a robot foundation model is trying to deploy is also the one an attacker wants to author. Call it the fifth bottleneck. What makes it different to the other four is that it tightens as the field matures: a more capable, more autonomous robot is a more dangerous compromised one.



The same capability also arms the offensive half of the security equation. Anthropic's private release of Claude Mythos Preview under Project Glasswing is a clear signal that frontier coding agents already have the potential of surfacing

UniPwn-class flaws in major operating systems.[12] The bar falls from both ends at once: intelligent software creates softer targets and sharper weapons.

Where this leaves us

Previously, we identified three key traits of a durable robotics business: proprietary data generation at scale, recurring revenue models, and low commoditization risk. Now, we add a fourth: **provable security**, which is the same as provable safety in physical AI. The commoditizing mass market is the worst place to be. The race to the bottom on cost is a race to the bottom on security, because cheap hardware and generalizable software create the largest attack surface there is. In physical AI, **security is not a feature that can simply be bolted on** by an OEM once regulation demands it, it is something that needs to be integrated into the supply chain from the very beginning.

A first wave of companies already helps OEMs do exactly that: Gray Swan, HiddenLayer, and Mindgard focus on the brain. On the body, Upstream Security and Trend Micro owned VicOne are extending their fleet-level security monitoring to physical AI, and Fort Robotics builds the safety-rated control platform that governs who can command a machine's actuators. But the layer that matters most: end-to-end guardrails governing the *robot's own* control authority remains almost entirely academic. This is a real whitespace opportunity waiting for someone to tackle it.

---

[1] South China Morning Post, Chinese researchers show how 1 word could allow spies to take control of a robot army, 2025

[2] Activant Research, Robots: Valuations and The Scaling Question, 2026

[3] Makris et al., Cybersecurity AI: Humanoid Robots as Attack Vectors, 2025

[4] Rest of World, The world's largest humanoid robot maker is going public, 2026

[5] Nottinghamshire Police, Officers testing 'revolutionary' robot dog, 2025

[6] CISA, ECOVACS DEEBOT Vacuum and Base Station (Update A), 2025

[7] AI Incident Database, Incident 842: Reportedly Hacked AI-Powered Robot Vacuums Allegedly Used for Surveillance and Harassment, 2024

[8] The New York Times, Waymo Suspends Service in Six Cities After Cars Drove Into Flooded Roads, 2026

[9] Zou et al., Universal and Transferable Adversarial Attacks on Aligned Language Models, 2026

[10] Jones et al., Adversarial Attacks on Robotic Vision Language Action Models, 2025

[11] Activant Research, Robots: Valuations and The Scaling Question, 2026

[12] Anthropic, Project Glasswing, 2026

Disclaimer: The information contained herein is provided for informational purposes only and should not be construed as investment advice. The opinions, views, forecasts, performance, estimates, etc. expressed herein are subject to change without notice. Certain statements contained herein reflect the subjective views and opinions of Activant. Past performance is not indicative of future results. No representation is made that any investment will or is likely to achieve its objectives. All investments involve risk and may result in loss. This newsletter does not constitute an offer to sell or a solicitation of an offer to buy any security. Activant does not provide tax or legal advice and you are encouraged to seek the advice of a tax or legal professional regarding your individual circumstances.

This content may not under any circumstances be relied upon when making a decision to invest in any fund or investment, including those managed by Activant. Certain information contained in here has been obtained from third-party sources, including from portfolio companies of funds managed by Activant. While taken from sources believed to be reliable, Activant has not independently verified such information and makes no representations about the current or enduring accuracy of the information or its appropriateness for a given situation.

Activant does not solicit or make its services available to the public. The content provided herein may include information regarding past and/or present portfolio companies or investments managed by Activant, its affiliates and/or personnel. References to specific companies are for illustrative purposes only and do not necessarily reflect Activant investments. It should not be assumed that investments made in the future will have similar characteristics. Please see "full list of investments" at activantcapital.com/companies/ for a full list of investments. Any portfolio companies discussed herein should not be assumed to have been profitable. Certain information herein constitutes "forward-looking statements." All forward-looking statements represent only the intent and belief of Activant as of the date such statements were made. None of Activant or any of its affiliates (i) assumes any responsibility for the accuracy and completeness of any forward-looking statements or (ii) undertakes any obligation to disseminate any updates or revisions to any forward-looking statement contained herein to reflect any change in their expectation with regard thereto or any change in events, conditions or circumstances on which any such statement is based. Due to various risks and uncertainties, actual events or results may differ materially from those reflected or contemplated in such forward-looking statements.