Robot domestici e cybersicurezza: i rischi reali da conoscere

I robot domestici basati su VLA non si limitano più a eseguire ordini, ma interpretano ambiente e linguaggio per decidere come agire. Per questo la cybersicurezza smette di essere un dettaglio tecnico e diventa una condizione essenziale della vita digitale in casa

Di Antonio Chella
Laboratorio di Robotica, dipartimento di Ingegneria Università degli Studi di Palermo 3/04/2026

I robot domestici basati su modelli Vision–Language–Action (VLA) segnano un passaggio decisivo: non eseguono soltanto istruzioni, ma osservano, interpretano e agiscono nello spazio quotidiano. È da qui che nasce un problema fondamentale: la cybersicurezza non riguarda più solo software e rete, ma l'intero processo che trasforma percezione e linguaggio in azione fisica.

Dal robot che esegue al robot che interpreta

In un mio precedente intervento sugli "umanoidi domestici" [1] sostenevo che la discontinuità più profonda nell'attuale ricerca robotica non è tanto la forma e la capacità motoria dei nuovi robot umanoidi, bensì il loro modo di apprendere e di interagire: il comportamento dei robot non è più basato sulla programmazione tradizionale, ma sull'apprendimento attraverso il linguaggio naturale e le dimostrazioni. Questa discontinuità è stata resa possibile, appunto, dai modelli Vision–Language–Action (VLA).

Oggi quel passaggio si sta consolidando e costringe a un nuovo cambio di prospettiva. Se un robot domestico diventa un agente che osserva, interpreta e agisce nello spazio più intimo della nostra vita quotidiana, allora la cybersicurezza non può essere un "requisito accessorio", ma diventa un requisito strutturale dell'abitare digitale.

Che cosa sono i robot domestici basati su VLA

Un robot basato su VLA può essere descritto, in modo sintetico, come un sistema robotico in cui l'azione dipende dallo stato interno del robot, da ciò che vede e da ciò che comprende attraverso il linguaggio. Il punto non è che il robot "capisce i comandi", ma che è capace di fondere la visione e il linguaggio in una rappresentazione comune dalla quale emergono le sue scelte operative e motorie. Inoltre, il robot è in grado di generalizzare dai dati di apprendimento. Questo paradigma è stato reso evidente da lavori fondativi come RT-2 di Google DeepMind, che integra modelli di linguaggio e di visione pre-addestrati su larga scala nel controllo robotico per migliorare la generalizzazione e il ragionamento [2].

Nel contesto domestico, il valore di questa generalizzazione è immediato: la casa è un ambiente non strutturato, pieno di eccezioni, e la possibilità di adattarsi con poche dimostrazioni è un requisito pratico. Anche qui la letteratura ha già mostrato segnali interessanti: il sistema Mobile ALOHA sviluppato alla Stanford University, ad esempio, evidenzia come l'imitazione da dimostrazioni raccolte tramite teleoperazione possa ottenere buone prestazioni su compiti domestici complessi, con un numero contenuto di esempi per ciascun compito [3].

Questa stessa architettura, però, sposta il "centro di gravità" della sicurezza: non basta più proteggere la rete Internet e il firmware. Bisogna proteggere l'intero processo che collega percezione, interpretazione e azione.

Perché i robot domestici basati su VLA ampliano la superficie d'attacco

Per capire perché la cybersicurezza cambia natura, conviene definire subito alcuni scenari plausibili in ambito domestico. Non sono "fantasie da hacker", ma conseguenze logiche del fatto che il robot prende decisioni in modo multimodale e produce output fisici.

Il linguaggio come canale di controllo

Il primo scenario riguarda il linguaggio come canale di controllo. Se la voce diventa un input operativo, allora la sicurezza non dipende solo da "cosa" viene detto, ma anche da "chi" lo dice e dalle proprietà fisiche del segnale. Esiste una letteratura consolidata che mostra come sistemi di riconoscimento vocale possano essere influenzati da comandi non udibili agli esseri umani, ma comunque interpretati dai dispositivi: DolphinAttack, ad esempio, dimostra la fattibilità di comandi veicolati in banda ultrasonica sfruttando la non-linearità dei microfoni, con effetti su assistenti vocali reali [4]. In una casa con robot in ascolto continuo, l'autenticazione dei comandi e la gestione del contesto

Robot domestici e cybersicurezza: i rischi reali da conoscere

sono questioni di sicurezza.

La prompt injection nel mondo fisico

Il secondo scenario riguarda la prompt injection, ossia la possibilità di deviare il comportamento di un sistema generativo tramite input formulati in modo opportunistico. La fondazione OWASP (Open Worldwide Application Security Project), nel suo Top 10 per le applicazioni LLM [5], pone la prompt injection tra i rischi principali, insieme al poisoning dei dati e alle vulnerabilità della supply chain. Nel caso dei robot basati su VLA, l'input linguistico non produce soltanto testo, ma influisce anche sulla scelta dell'azione. Inoltre, in casa l'input linguistico può provenire non solo dalla voce dell'utente, ma anche da testi presenti nell'ambiente (come scritte, schermi, istruzioni contestuali). Il mondo fisico smette di essere "sfondo" e diventa un canale di input. Si pensi, ad esempio, a una scritta lasciata su un pezzo di carta che fornisce un'istruzione malevola al robot.

Gli attacchi percettivi e le decisioni errate

Il terzo scenario riguarda gli attacchi percettivi. Sappiamo da tempo che i modelli di visione possono essere influenzati in modo robusto nel mondo reale tramite schemi stampabili. Il lavoro su Adversarial Patch [6] mostra patch universali e fisicamente realizzabili che, pur non essendo "magia", spostano sistematicamente l'output di classificatori anche in presenza di trasformazioni realistiche (rotazioni, scale, illuminazione). Nel robot domestico, la conseguenza non è solo un'etichetta errata, ma può anche essere una decisione errata. Se il robot interpreta in modo scorretto un oggetto o una scena, l'errore può tradursi in presa, movimento, manipolazione.

La supply chain dei modelli e del software

Il quarto scenario, spesso sottovalutato nel dibattito pubblico, è quello della filiera dei modelli e del software. I robot domestici basati su VLA raramente partono da zero: usano componenti pre-addestrati e tecniche di fine-tuning. Questo rende concettualmente plausibili attacchi "a monte", in cui il comportamento anomalo non deriva da una compromissione della rete domestica, bensì da una backdoor inserita nella supply chain del modello. BadNets [7] è un riferimento classico perché mostra come un modello possa funzionare "normalmente" sui dati puliti e attivare comportamenti deviati in presenza di un evento specifico. Anche la fondazione OWASP include esplicitamente il data poisoning e le vulnerabilità della supply chain tra i rischi sistemici.

Gli aggiornamenti come parte della sicurezza

Il quinto scenario riguarda gli aggiornamenti. In un robot domestico che evolve nel tempo, un aggiornamento non è solo manutenzione, ma può cambiare il comportamento del robot. Questo rende fondamentali la necessità di tracciare le versioni, la verificabilità degli aggiornamenti, la possibilità di tornare alle versioni precedenti e la verifica delle modifiche apportate. Il punto non è "diffidare degli aggiornamenti", ma progettare un ciclo di vita in cui l'evoluzione del sistema di controllo del robot sia gestita come parte della sicurezza.

L'ecosistema IoT e il movimento laterale

Il sesto scenario nasce dall'ecosistema IoT domestico. Il robot non vive in isolamento: convive con router, altoparlanti, telecamere, serrature, sensori. In caso di compromissione, può diventare un punto di osservazione per il movimento laterale verso altri dispositivi, poiché è spesso il nodo con più sensori e maggiore capacità di calcolo. Qui diventano rilevanti le normative tecniche per la sicurezza IoT consumer, come ETSI EN 303 645 [8], e le linee guida come la NISTIR 8259A [9], che formalizzano requisiti minimi di sicurezza e capacità di gestione lungo il ciclo di vita dei dispositivi connessi. Se, inoltre, il robot utilizza middleware robotici, la sicurezza delle comunicazioni interne assume un ruolo diretto: la documentazione ROS 2 descrive l'uso di DDS-Security tramite sros2 per proteggere le comunicazioni tra nodi [10].

Quando l'indisponibilità diventa rischio fisico

Infine, anche l'indisponibilità può diventare un rischio fisico. La fondazione OWASP include il model denial of service tra i rischi degli LLM perché gli attacchi di esaurimento delle risorse possono degradare o interrompere il servizio. In un robot, la degradazione non è solo "un servizio lento": può comportare latenza decisionale, fallback non ottimali e una perdita di fluidità

Robot domestici e cybersicurezza: i rischi reali da conoscere

del controllo in situazioni in cui la reattività è parte della sicurezza.

Dal cyber al safety: quando un bit diventa un newton

Questi scenari chiariscono un punto: nei robot domestici basati su VLA la separazione storica tra security (attacchi intenzionali) e safety (guasti accidentali) tende a collassarsi. Un attacco non deve "rompere" il sistema; può spingerlo verso decisioni plausibili ma sbagliate. E l'effetto finale non è solo digitale: la forza, il movimento, l'interazione del robot.

Per questo la cybersicurezza dei robot domestici non può essere confinata al perimetro informatico. Deve essere un disegno multilivello che attraversa l'identità del dispositivo, l'integrità della catena, la sicurezza delle comunicazioni, la robustezza del modello, i vincoli fisici certificati e i guardrail semantici. Questi ultimi sono particolarmente importanti: un sistema che interpreta il linguaggio naturale deve riconoscere le ambiguità, chiedere conferma quando necessario e rifiutare azioni ad alto rischio se il contesto non è sufficientemente chiaro. In altri termini, la "sicurezza semantica" diventa complementare alla sicurezza meccanica.

On-device e continuità operativa: una scelta tecnica che diventa scelta di governance

Nella prospettiva domestica, l'esecuzione locale di parti rilevanti dell'architettura cognitiva è anche una leva per la sicurezza e la privacy. Riduce la dipendenza dalla rete, migliora continuità operativa e limita l'esposizione dei dati sensibili della vita domestica.

In questa direzione vanno iniziative industriali recenti: Google DeepMind ha introdotto Gemini Robotics [11] come famiglia di modelli per integrare capacità multimodali nel controllo fisico e ha presentato una variante on-device progettata per l'inferenza locale e per l'adattamento rapido ai compiti. Il punto, però, non è "cloud sì o cloud no", ma il governo di un'architettura ibrida: ciò che può restare in casa dovrebbe restare in casa, e ciò che deve passare sul cloud dovrebbe farlo in modo verificabile, minimo e tracciabile.

Politica europea: AI Act e Cyber Resilience Act rendono la sicurezza un requisito di prodotto

Qui entra in gioco il quadro europeo. L'AI Act [12], nella pagina ufficiale della Commissione, prevede un'applicazione progressiva: pratiche proibite e obblighi di alfabetizzazione sull'AI dal 2 febbraio 2025, obblighi per i modelli di AI per finalità generali e governance dal 2 agosto 2025, piena applicabilità dal 2 agosto 2026, con un periodo di transizione fino al 2 agosto 2027 per alcune regole legate ai sistemi ad alto rischio incorporati in prodotti regolati. Per la robotica domestica basata su VLA, questo significa che robustezza e cybersecurity non sono più soltanto "best practice", ma aspetti che entrano nel perimetro di conformità e responsabilità, soprattutto quando il robot opera in scenari che incidono sulla sicurezza delle persone o sull'assistenza a soggetti fragili.

Il Cyber Resilience Act [13], dal canto suo, sposta la cybersecurity nella logica della sicurezza di prodotto, con marcatura CE e obblighi lungo il ciclo di vita. La Commissione indica che il Cyber Resilience Act è entrato in vigore il 10 dicembre 2024; le principali obbligazioni si applicano dall'11 dicembre 2027 e gli obblighi di reporting dal 11 settembre 2026. Per i robot domestici con update continui e dipendenze complesse, questo implica processi industriali maturi per la gestione delle vulnerabilità, la tracciabilità e gli aggiornamenti: un requisito che, di fatto, diventa parte dell'ingegneria del robot tanto quanto i sensori e gli attuatori.

Anche la direttiva NIS2 [14], con la scadenza di recepimento fissata al 17 ottobre 2024, indica la direzione europea: la governance del rischio, la supply chain e le capacità di enforcement sono centrali e contribuiscono a un ecosistema in cui l'insicurezza di filiera è sempre meno tollerata.

Verso un'ingegneria della sicurezza cognitiva

La diffusione dei robot intelligenti in casa non dipenderà solo dalla destrezza o dall'abilità di dialogare. Dipenderà dalla fiducia. E la fiducia, nel digitale, è un costrutto tecnico e istituzionale insieme: nasce dalla trasparenza, dalla verificabilità, dalla gestione responsabile degli aggiornamenti e dalla capacità di mitigare rischi prevedibili.

In termini di metodo, un riferimento utile per collegare rischi, contesto d'uso e misure organizzative e tecniche

Robot domestici e cybersicurezza: i rischi reali da conoscere

è il NIST AI Risk Management Framework [15], che propone un approccio sistematico alla gestione del rischio AI senza ridurlo a un singolo livello (modello, dato o deployment). Nel caso dei robot basati su VLA, questo approccio conduce a una conclusione semplice: la cybersicurezza deve proteggere non soltanto il robot come dispositivo, ma anche il robot come decisore con un corpo. È qui che serve un'"ingegneria della sicurezza cognitiva" capace di integrare la sicurezza informatica classica, la robustezza dei modelli, i vincoli fisici e la governance degli aggiornamenti.

Perché la cybersicurezza dei robot domestici basati su VLA è un prerequisito

Se vogliamo che i robot domestici basati su VLA diventino una tecnologia quotidiana e non una curiosità da demo, dobbiamo considerare la cybersicurezza come prerequisito dell'abitazione digitale. In un sistema che vede, interpreta e agisce, la protezione riguarda anche il perimetro cognitivo, perché è lì che un input malevolo, ambiguo o fuori distribuzione può trasformarsi in un comportamento fisico. Il quadro europeo, dall'AI Act al Cyber Resilience Act, spinge nella direzione giusta: non per frenare l'innovazione, ma per renderla scalabile e affidabile, imponendo responsabilità lungo la filiera e la gestione del rischio lungo il ciclo di vita. La partita, nei prossimi anni, non sarà decidere se i robot entreranno in casa, ma con quale architettura: modelli più robusti, aggiornamenti verificabili, dati realmente governati e comportamenti prevedibili quando il contesto diventa incerto. Solo così l'intelligenza incorporata nel robot potrà diventare un'alleata dell'autonomia domestica senza trasformarsi in una nuova fonte di fragilità.

Bibliografia

[1] A. Chella, Robot umanoidi domestici: la rivoluzione AI entra in casa, Agenda Digitale 2025 https://www.agendadigitale.eu/cultura-digitale/robot-umanoidi-domestici-la-rivoluzione-ai-entra-in-casa/

[2] A. Brohan et al., "RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control", arXiv:2307.15818.

[3] Z. Fu, T. Z. Zhao, C. Finn, "Mobile ALOHA: Learning Bimanual Mobile Manipulation with Low-Cost Whole-Body Teleoperation", arXiv:2401.02117.

[4] G. Zhang et al., "DolphinAttack: Inaudible Voice Commands", arXiv:1708.09537.

[5] https://owasp.org/www-project-top-10-for-large-language-model-applications/

[6] T. B. Brown et al., "Adversarial Patch", arXiv:1712.09665.

[7] T. Gu et al., "BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain", arXiv:1708.06733.

[8]https://www.etsi.org/deliver/etsi_en/303600_303699/303645/03.01.03_60/en_303645v030103p.pdf

[9] https://doi.org/10.6028/NIST.IR.8259A

[10] https://docs.ros.org/en/kilted/Tutorials/Advanced/Security/Introducing-ros2-security.html

[11] https://deepmind.google/models/gemini-robotics/

[12] https://digital-strategy.ec.europa.eu/it/policies/regulatory-framework-ai

[13]https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act

[14]https://digital-strategy.ec.europa.eu/it/policies/nis2-directive

[15] https://www.nist.gov/itl/ai-risk-management-framework