



ORDINE DEGLI INGEGNERI DI ROMA - FOIR - LEZIONE 6

COD. 68 - CORSO AVANZATO SULL'INTELLIGENZA ARTIFICIALE
11 APRILE 2026

NUOVI DEVICES ED UMANOIDI

Dai Large Language Model (LLM) ai Large Action Model (LAM)

Architettura VLA · Limiti Hardware · Regolamento Macchine · ROI



**Ing.
Fabrizio
Catinari**
Ingegnere
Professionista
INGEGNERIA E
INNOVAZIONE



Perché l'Umanoide?

La sfida tra due filosofie produttive e infrastrutturali opposte.

Automazione Tradizionale

Richiede ambienti **Greenfield** (costruire un impianto da zero o modificarlo pesantemente). Bracci antropomorfi o SCARA esigono che la linea sia progettata *attorno* al robot. Necessitano di pavimenti industriali perfetti, gabbie di sicurezza in acciaio e dime posizionate al millimetro.

Form Factor Umano

Progettato per il **Brownfield**. Il nostro mondo è costruito per bipedi alti 170 cm. L'umanoide usa le stesse scale degli operai, le stesse porte, e impugna trapani progettati per mani umane. Azzera il CapEx infrastrutturale.



VLA: vedere → capire → agire

Il VLA è un LAM specializzato che genera direttamente comandi motori in loop sensoriale.



Il VLA non genera testo o API call,
genera angoli di giunto e vettori di
forza.

**Questo è il salto dalla
semantica pura alla
cinematica differenziale.**



LLM vs LAM: L'Action Space

Come pensa un modello generativo applicato alla fisica vettoriale?

Language Model (LLM)

Lo spazio probabilistico è il vocabolario testuale. Il modello prevede statisticamente la parola (Token) successiva.

Input: "Il robot si..."

Output Token Softmax:
[muove] (85%) | [ferma] (15%)

Action Model (LAM)

Lo spazio continuo (XYZ, Roll, Pitch, Yaw) viene Quantizzato (Voxel). Il modello prevede la posa (Waypoint vettoriale) successiva.

Generative Action Token:

[Δx :+2, Δy :-1, Δz :0, Torque:14Nm]



Percezione Spaziale: Dal 2D al 3D

I sensori hardware che alimentano il cervello VLM.

1. Visione Stereoscopica

Usa due lenti calibrate per calcolare la *disparità* dei pixel. Calcolo computazionalmente leggero, ma fallisce al buio o su lamiere lisce e specchianti tipiche dell'automotive.

2. Depth / ToF

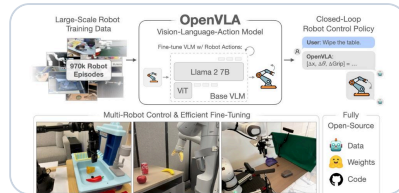
Sensori attivi (IR). Non dipendono dalla luce ambientale. Altissima risoluzione per la micro-manipolazione e il grasping ad anello chiuso (< 2 metri).

3. Sensori LiDAR

Laser Solid-State. Genera una Point Cloud (nuvola di punti) ad alta frequenza. Fondamentale per la macro-navigazione SLAM in fabbrica fino a 100m, bypassando polvere o riflessi.



Modelli VLA — Lo stato dell'arte



OpenVLA

OPEN-SOURCE

UC BERKELEY

7B param, base accademica di riferimento per manipolazione spaziale fine e applicazioni On-Premise (Air Gap).

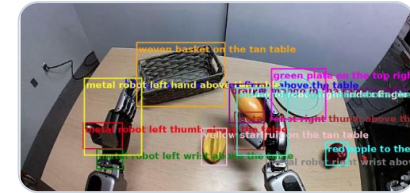


UnifoLM-VLA

OPEN-SOURCE

UNITREE ROBOTICS

Addestrato intensivamente sui dati di teleoperazione dell'umanoide Unitree G1. Ottimo entry-level industriale.

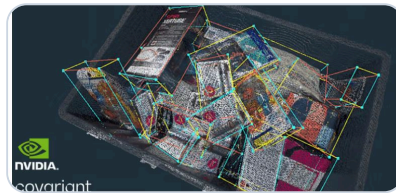


GROOT N1.5

PARTNER

NVIDIA

Ecosistema profondamente integrato con i motori fisici in ray-tracing per la generazione di dati sintetici (Isaac Sim).

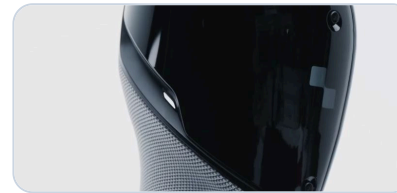


RF-1 (Robotics FM)

CHIUSO

COVARIANT

Modello "Video-Language-Action" nato per l'industria. Addestrato su milioni di reali azioni logistico-robotiche.

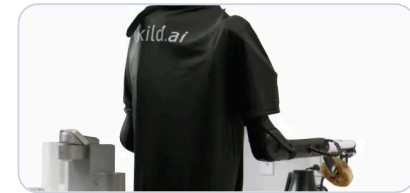


Helix

CHIUSO

FIGURE AI

Architettura neurale "Fast/Slow" per incorporare le decisioni tattiche logiche da quelle cinematiche real-time.



Skild Brain

CHIUSO

SKILD AI

Foundation Model adattivo. Progettato per scalare dall'automazione logistica fino agli ambienti non strutturati operando in Zero-Shot su hardware eterogenei.



Il grande collo di bottiglia: I Dati

Non esiste un "internet delle azioni fisiche" da cui fare scraping massivo.

Teleoperazione Umana

Operatori umani indossano visori e **guanti aptici** per pilotare il robot in prima persona e registrare vettorialmente migliaia di traiettorie ottimali. Processo ingegneristico lento e ad alto costo (Imitation Learning).



Dati Sintetici Massivi (Sim)

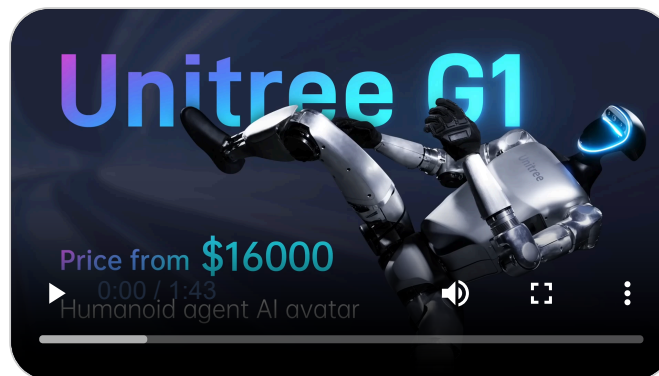
Si usano motori fisici in real-time raytracing per generare digital twin 3D in cui replicare il task milioni di volte variando luci, masse e attriti (Domain Randomization).



DINAMICA E CONTROLLO HW

Unitree G1 vs G1-Spring

MODELLO BASE (LENTO / IK)



Cinematica Inversa tradizionale. Movimento controllato ma fragile.

G1 SPRING (VELOCE / RL POLICY)



Cinematica via Reti Neurali. Il LAM ottimizza la dinamica globale (Momento Angolare), permettendo recovery reattivo istantaneo.



Panorama Umanoidi (2026)

Azienda	Robot	Status 2026	Note Commerciali
Unitree CN	G1, H1	Produzione di massa	Domina il mercato per il costo di ingresso irrisorio, grazie all'integrazione verticale della BOM hardware (~\$16k per il G1).
Figure AI US	Figure 02	Deploy OEM	Pilota industriale validato presso le linee di assemblaggio scocche in BMW (Spartanburg).
Tesla us	Optimus Gen3	Uso interno (Beta)	Forte verticalizzazione degli attuatori lineari, ma ancora assenti clienti terzi (B2B).
Agility us	Digit	Logistica RaaS	Modello puramente ad noleggio RaaS in test logistici intensivi (GXO, Amazon).



R&D vs Prodotto: Il punto critico



Cosa NON mostrano i video virali

- Il **Success Rate reale** prima della ripresa perfetta (Cherry Picking commerciale).
- Teleoperazione nascosta (Tecnica "Wizard of Oz") spacciata per autonomia completa della rete neurale.
- **Tempi di ciclo (Takt Time)**: Attualmente 3-5 volte più lenti dei ritmi imposti a un operatore umano standard.



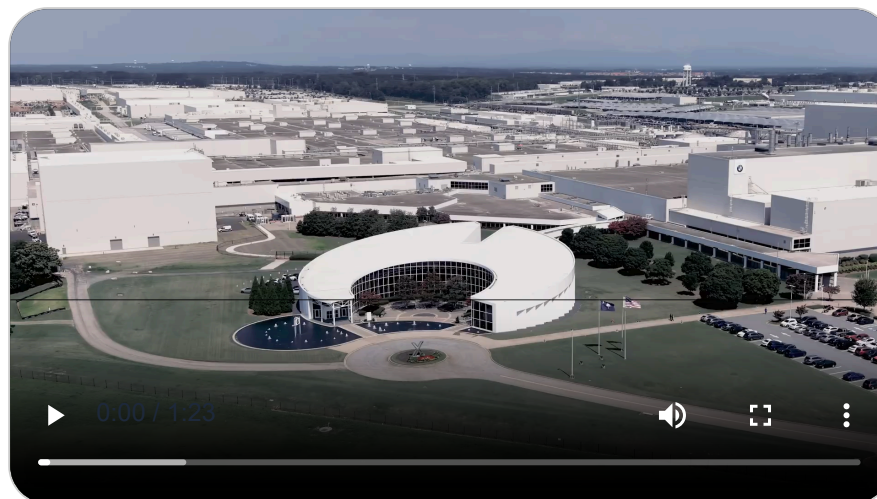
I traguardi industriali effettivi

- Figure 02 ha dimostrato turni operativi in linea BMW interfacciandosi con successo ai sistemi SCADA di impianto.
- Consolidamento dei motori fisici hardware-accelerated (Nvidia Isaac) per arginare il "Sim-to-Real gap".
- Modelli VLA Open Source testabili offline dalle PMI tutelando il segreto industriale dei propri semilavorati.



AMBIENTE STRUTTURATO

Figure 02 in BMW



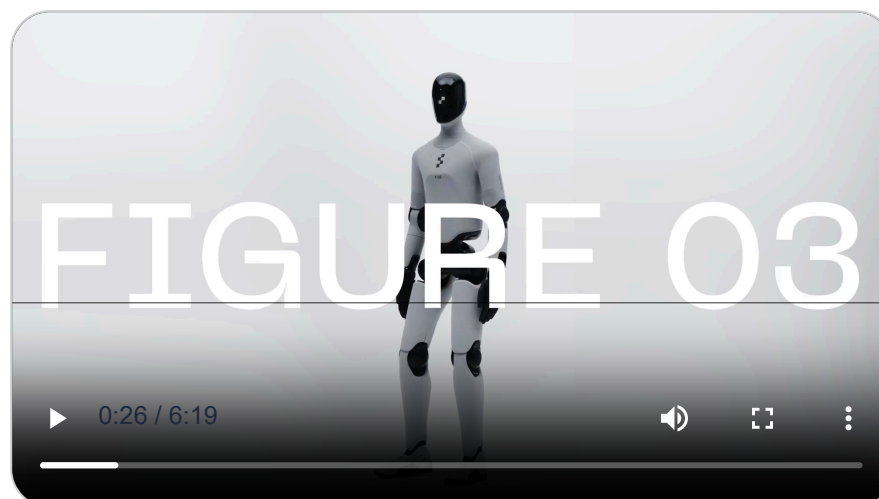
Integrazione in fabbrica reale.
L'operazione di assemblaggio ad alta tolleranza (**Pin-in-hole** su lamiere rigide) dimostra la bontà del controllo di impedenza attiva sui giunti per evitare micro-danni alla carrozzeria.

- ✓ Interfaccia logica I/O con sistemi MES aziendali
- ✓ Navigazione fusa con AGV/AMR logistici
- ✓ Resistenza a stress termico e particolato metallico



AMBIENTE CAOTICO NON STRUTTURATO

Figure 03: Casa



Il test estremo. L'ambiente domestico richiede un adattamento "Zero-Shot". Non esistono CAD o planimetrie preventive. L'hardware affronta ostacoli deformabili (cuscini, tappeti) e imprevisti continui in presenza di bambini/animali.

- ✓ Segmentazione Semantica 3D generativa real-time
- ✓ Gestione sicura delle collisioni soffici
- ✓ NLP (Natural Language) Human-in-the-loop fluido



I veri colli di bottiglia hardware

1. Edge Compute vs Cloud

L'inferenza LAM pesante deve avvenire su **NPU locali in edge**. Mandare flussi video via API Cloud Wi-Fi introduce latenze di rete catastrofiche (Jitter > 100ms) che destabilizzano gli algoritmi MPC, facendo letteralmente cadere a terra il robot.

2. Sensori Tattili Elastomerici

La micro-manipolazione industriale esige sensibilità. Quando il palmo copre visivamente l'oggetto (Self-Occlusion), le camere sono inutili. Servono **pelli elettro-capacitive** per ricalcolare lo "Slip Vector".

3. Il "Sim-to-Real" Gap

Un LAM addestrato al PC ignora l'usura delle cinghie, il surriscaldamento termico degli avvolgimenti dei motori (che abbassa la resa di Coppia) o i giochi meccanici (Backlash). Il divario va colmato con pesanti randomizzazioni.



~1.000

morti sul lavoro/anno in Italia

~600k

infortuni denunciati nel 2024

40%

degli incidenti gravi in industria

#1

causa: cadute e "Spazi Confinati"

Ispezione Gallerie

⚠ RISCHIO ASFISSIA, CROLLO STRUTTURALE

Analisi ultrasonora in tunnel ferroviari o condotte forzate senza sistemi di aerazione attivi.

Spazi ATEX Confinati

⚠ RISCHIO INNESCO ESPLOSIONE VAPORI

Manutenzione serbatoi e reattori chimici. Il "Zero-Entry" elimina radicalmente la burocrazia dei permessi D.Lgs. 81/08.

Miniere Sotterranee TRL 5

⚠ RISCHIO CEDIMENTO STRUTTURALE

Teleoperazione estrattiva in zone profonde instabili (es. miniere Harmony Gold in Sudafrica).

Nucleare & Decommissioning

⚠ ESPOSIZIONE RADIAZIONI IONIZZANTI

Remote handling per lo smantellamento vasche di reattori (ispezione a distanza sicura e manipolazione residui).

Impianti O&G (Raffinerie)

⚠ SHOCK TERMICO (SUPERFICI >300°C)

Lettura valvole e clampaggio tubi in marcia, permettendo diagnosi senza fermare le torri di distillazione.

Cantieri Stradali Notturni

⚠ INVESTIMENTO LETALE SU CORSIE

Robot adibiti al posizionamento automatico di segnaletica (coni) a bordo autostrada su mezzi pesanti sfreccianti.



La Sicurezza dell'Interazione (HRI)

Intrinsecamente Sicuri e PFL

I giunti usano motori a bassa inerzia riflessa integrando misuratori di coppia precisissimi. All'urto anomalo con l'uomo, i motori "cedono" istantaneamente entrando in modalità backdrivable (Compliance attiva/passiva), dissipando l'energia meccanica dell'impatto.

LiDAR + VLM Boundaries

Le gabbie protettive in metallo spariscono dalla linea. La fusione tra LiDAR spaziale e telecamere VLM crea recinti logici virtuali (SSM). Se l'operatore invade la zona, l'IA rampa dolcemente la velocità cinematica verso lo zero, mantenendo un assetto statico sicuro per la macchina.



Certificazione e Compliance: 3 Strati

Regolamento Macchine (UE) 2023/1230

Trasforma il Machine Learning in un componente critico certificabile. Impone il congelamento del "Continual Learning" libero in ambiente di produzione (peso sinaptico bloccato o *Frozen Weights*) per garantire prevedibilità.

EU AI Act (2024) — L'Obbligo Human-in-the-loop

Un robot pilotato da un VLA = Sistema AI ad **Alto Rischio**. Esige registrazione documentale dei Dataset di training, logging inalterabile (tipo Scatola Nera) e chiara interfaccia per l'override umano istantaneo.

Cyber Resilience Act & ISO 10218

Impone rigorosi standard di Cyber-Security by Design. Requisito cruciale: **Isolamento Elettromeccanico (Air Gap)** tra il circuito elettrico dei motoriduttori di potenza (Safety PLC) e le logiche cloud-based della NPU.



Il robot VLA sbaglia: Chi risponde?

Analisi delle responsabilità legali se un Large Action Model allucina un ostacolo e innesca un impatto critico.

1. Provider (Fabbricante AI)

Sviluppa l'architettura dei Foundation Models e l'hardware base. Ha l'onere legale di dichiarare i limiti operativi, il Bias residuo dei dataset e i test di sicurezza documentati. Risponde legalmente per i bug nativi, crash architetturali o violazioni di privacy dei dati visivi acquisiti.

2. Integrator (Noi Ingegneri)

Concepisce e progetta la Cella Robotizzata. Esegue, valida e firma la Marcatura CE e l'Analisi Rischi obbligatoria (ISO 12100). Determina le posizioni degli scanner e i limiti cinematici di Coppia/Velocità sui drive dell'umanoide.

3. Deployer (Azienda Cliente)

Società utilizzatrice finale ed esercente dell'impianto. Ha il dovere di formare adeguatamente gli addetti di linea esposti (D.Lgs. 81/08). Risponde in solido se aggira i sistemi HRI (es. bypass di sicurezza) o permette un utilizzo difforme dai manuali rilasciati dall'Integratore.



Paradigma Economico (Il ROI Industriale)

La Rivoluzione dei Costi Strutturali (CapEx)

- **Dal Custom allo Standard:** Una cella protetta per il kitting randomico con braccio fisso a sei assi costa in media chiavi in mano ~150.000€.
- **Payload vs Agilità Spaziale:** L'umanoide cede volontariamente la ripetibilità del centesimo di millimetro (tipica dello SCARA) in favore di un'adattabilità generale e reattiva "human-like".
- **No Hard-Coding Oneroso:** Riprogrammare le routine di uno SCARA richiede costose giornate di consulenza specializzata PLC. L'architettura LAM promette l'istruzione via linguaggio naturale (NLP Zero-Code).

**Da
~150k€**

Costo Medio Cella Automazione Fissa Integrata



a ~16k€

Hardware Umanoide Base Standard (Unitree G1)



LIVE IT ARCHITECTURE

La Pipeline in Pratica

Un'interfaccia basata su paradigma **Zero-Code**.
Drastica riduzione della curva di apprendimento
per il personale non laureato in azienda.

1. **HRI/LLM**: Il modello processa il "Natural Prompt" estratto dal tecnico.
2. **VLM Pipeline**: Analisi spaziale del Base64 snapshot proveniente dalla fotocamera della cella.
3. **VLA Parser**: Genera l'Action Vector in formato JSON strutturato o nativo G-Code per l'esecuzione cinematica.

AVVIA DIGITAL TWIN LIVE →

```
root@foir-server:/backend/proxy_lam.php

// 1. Acquisizione Context Multi-Modale (WebGL + HRI)
$human_prompt = $_POST['prompt']; // "Seleziona cubo rosso"
$camera_vision = $_POST['base64_image']; // Array di Pixel Snapshot 2D

// 2. Trasmissione al VLA Tensor Inference Engine (Secure API HTTP)
$tensor_response = curl_exec($ch_secure_connection);

// 3. Parsing del Generative Output in Waypoint Matematici
{
  "kinematics_target": { "x": 1.57, "y": 0.8, "z": -0.5
},
  "rotation_quaternion": [0, 0.707, 0, 0.707],
  "gripper_state": "engage_close_loop"
}
```



Digital Twin: Live Simulator

VLA Zero-Code Interaction Framework

Fullscreen Mode

R

Embodied AI Core
SOCIAL HRI + VLA | ING. FABRIZIO CATINARI

MOTORE LAM

Llama 4 Scout 17B (Groq)

POST-PROCESSOR HW

Standard G-Code

VLA EXPERT MODE

Invia snapshot reale della camera VLM al LLM

☐

1. VLM

→

2. LAM

→

3. ROS

→

4. HW





Alta Formazione In-House FOIR

Supportiamo e addestriamo i board tecnici aziendali per integrare la robotica generativa e IA curando contestualmente il rigoroso stile **Corporate**.

Consulta il programma su aziende.foir.it



Q&A Engineering

Spazio al dibattito tecnico

Validazione safety-critical di architetture VLA e mitigazione del rischio CE (Reg. 2023/1230) per la transizione da cinematica deterministica a reti neurali in ambienti Brownfield. Se stai affrontando questi nodi architetturali e hai voglia di confrontarti, parliamone:



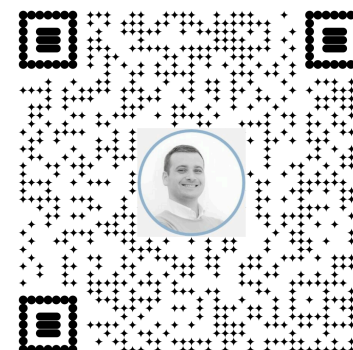
Ing. Fabrizio Catinari

Ingegneria e Innovazione

+39 335 7898179

fabrizio@ingcatinari.com

www.ingcatinari.com



 LinkedIn