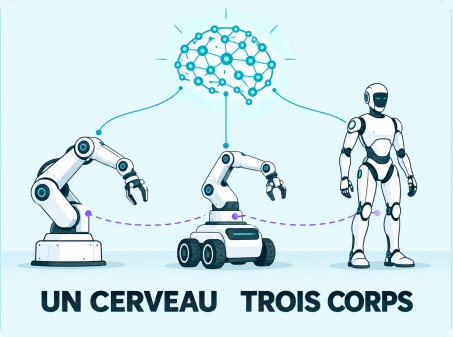
# VLA
# un cerveau pour plusieurs corps, entre démonstrations et robots utiles




0.7, GEMINI ROBOTICS 2, HELIX 02, GR00T N1.7, MOLMOACT 2 : DES MODÈLES QUI COMMANDENT UN CORPS ENTIER, UN GOULET QUI PASSE AUX DONNÉES (RELEVÉ AU 18 SEPTEMBRE 2026)

Un modèle Vision-Language-Action regarde, comprend une consigne en langage naturel et produit directement des commandes pour un robot. En 2026, ces « cerveaux » généralistes pilotent bras, mains et jambes, et l'un d'eux tourne déjà dans une usine BMW. Mais entre la démonstration et le robot utile à grande échelle, le goulet est passé du modèle aux données, à la fiabilité et à la sécurité : l'horizon se compte en années, pas en mois. Cette page explique le principe, passe en revue les acteurs fermés et ouverts, confronte démonstrations et déploiements, nomme les verrous, puis classe chaque promesse selon son degré de certitude.

ROBOTIQUE | HUMANOÏDES | DONNÉES | SÉCURITÉ

## » REPÈRES SEPTEMBRE 2026

**5 500+**
HUMANOÏDES LIVRÉS PAR LE N°1 MONDIAL EN 2025
Unitree, 32 % du marché selon son propre prospectus ; entrée en Bourse à Shanghai le 19 août 2026.

**20 854 h**
DE VIDÉO HUMAINE DANS UN VLA OUVERT
Vidéo égocentrique (EgoScale) pour GR00T N1.7, avril 2026 : NVIDIA y voit une « loi d'échelle de la dextérité ».

**30 000+**
BMW X3 PRODUITES AVEC LA CONTRIBUTION DE FIGURE 02
En environ 10 mois ; Figure 03 et Helix 02 sont entrés en production à Spartanburg en juin 2026.

**18 h**
DE PRÉPARATION D'ESPRESSO SANS INTERRUPTION
Mesure publiée par Physical Intelligence pour À*0.6r(ovembre 2025), après apprentissage par l'expérience.

## » SOMMAIRE

01 Un modèle qui voit, comprend et agit
02 Qui construit quoi, en septembre 2026
03 Démonstrations contre déploiements
04 Les verrous : données, fiabilité, temps réel, sécurité
05 L'horizon : de l'acquis au spéculatif
06 Que faire de tout cela en 2026

PRINCIPE / 01

01

# Un modèle qui voit, comprend et agit

Un VLA ne produit pas du texte mais des actions, des dizaines de fois par seconde. Trois idées expliquent pourquoi c'est difficile et comment on le contourne : produire des mouvements continus, prédire par paquets, séparer le réflexe de la réflexion.

**Entrée : des pixels et une consigne. Sortie : des actions du robot.** Le terme naît en 2023 avec RT-2 (Google DeepMind) ; en 2026, presque chaque grand laboratoire d'IA et plusieurs start-up très financées ont leur modèle de fondation robotique.

01 DÉFINITION

### Il voit, comprend et agit

À partir d'un flux caméra et d'une instruction (« range la vaisselle »), le modèle émet des commandes : positions d'articulations, déplacement de la main, ouverture de la pince.Un modèle vision-langage classique répond par du texte ; un VLA **pilote un corps réel**, en continu.

02 APPRENTISSAGE

### Il hérite du bon sens d'internet

Son socle est un modèle vision-langage pré-entraîné : il sait ce qu'est une tasse ou « ranger », même sur un objet jamais vu en démonstration. À0.6associe ainsi un modèle de 5 Md de paramètres à un « action expert ».Puis il apprend à agir par **imitation** : téléopération, vidéo humaine, simulation.

03 CONTEXTE

### Il est le voisin des modèles de monde

Un modèle de monde prédit ce qui va se passer ; un VLA sort directement l'action. Les deux convergent : Cosmos 3 (NVIDIA) prédit déjà l'action et sa conséquence visuelle.Voir la page **World Models**, qui traite l'autre moitié du sujet.

» SOUS LE CAPOT : QUATRE CHOIX D'INGÉNIERIE

Un robot vit dans un monde continu, en temps réel : une décision qui arrive 200 ms trop tard fait échouer la tâche. D'où des solutions très différentes de celles d'un modèle de langage.

01

### Des mouvements continus : le flow matching

Les premiers VLA (RT-2, OpenVLA) découpaient les actions en « tokens », comme des mots : simple, mais saccadé et mauvais sur les tâches de contact (plier du tissu). **À (octobre 2024)** génère des trajectoires continues par flow matching, un procédé proche de la diffusion ; SmolVLA et d'autres l'ont adopté. MolmoAct raisonne d'abord en 3D, via des tokens de profondeur.

02

### Des paquets d'actions : l'action chunking

Un bras se commande vers 50 Hz (une commande toutes les 20 ms), ce qu'aucun grand modèle ne tient. Il prédit donc un **chunk de 10 à 50 actions** d'un coup (50 actions, soit environ 1 s, pour À0). Le Real-Time Chunking (9 juin 2025) rend l'inférence asynchrone et tolère des délais de plus de 300 ms. Mesure du papier À0 (2024) sur RTX 4090 : 46 ms de préremplissage, 73 ms par chunk.

03

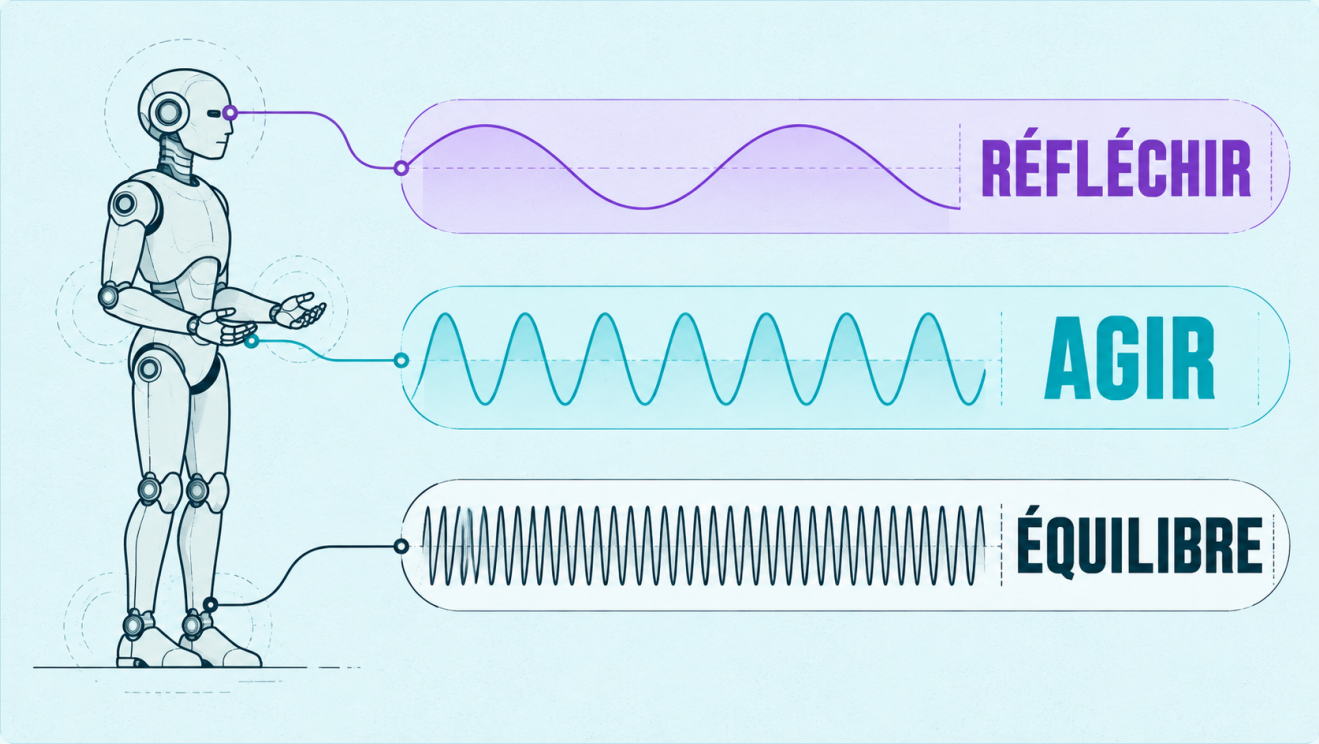
### Trois vitesses : le réflexe et la réflexion

**Helix (Figure, février 2025)** sépare un modèle lent (7 Md de paramètres, 7 à 9 Hz) d'une politique rapide (80 M, 200 Hz). **Helix 02 (27 janvier 2026)** commande le corps entier avec trois niveaux : S2 lent, S1 à 200 Hz, et S0 à 1 kHz pour l'équilibre et les contacts, qui remplace 109 504 lignes de C++.GR00T de NVIDIA suit la même logique à double système.

04

### Un cerveau, plusieurs corps

Entraîner un seul modèle sur plusieurs robots améliore **souvent** les résultats, surtout pour les robots qui disposent de peu de données. Open X-Embodiment (octobre 2023) rassemble plus d'un million de trajectoires, 22 types de robots, 21 institutions : +50 % en régime de peu de données pour RT-1-X.Le gain n'est pas systématique, et l'hétérogénéité des espaces d'action reste un verrou de recherche.




Les trois vitesses du contrôle : réfléchir (quelques fois par seconde), agir (200 fois), tenir l'équilibre (1 000 fois).

ACTEURS / 02

02

# Qui construit quoi, en septembre 2026

Trois cerveaux fermés, dont deux de laboratoires de pointe, puis six repères du côté ouvert et européen, ceux que l'on peut télécharger et faire tourner soi-même.

**Trois acteurs fixent le rythme, tous propriétaires.** Leurs comparaisons de performance sont celles des éditeurs eux-mêmes, sans évaluation indépendante.

01 PROPRIÉTAIRE

### Physical Intelligence · À0.7

À*0.6 (17 novembre 2025) ajoute l'apprentissage par renforcement à l'imitation. **À0.7 (16 avril 2026)** recombine des savoir-faire appris séparément pour des tâches jamais vues (une friteuse à air, guidée pas à pas) et, selon l'éditeur, égale des modèles spécialisés.Limites reconnues : il ne sait pas exécuter seul une consigne de haut niveau, et la réussite dépend de la formulation (de 5 % à 95 % sur une tâche). Seuls À0, À0-FAST et À0.5 sont ouverts.

02 ACCÈS RESTREINT

### Google DeepMind · Gemini Robotics 2

**30 juillet 2026** : trois modèles. Un VLA qui contrôle tout le corps d'un humanoïde (démonstration sur l'Apollo 2 d'Apptronik), ER 2 (raisonnement, orchestration de plusieurs robots) et On-Device 2, qui s'adapte à un nouveau robot en quelques heures avec moins de 200 exemples.Seul ER 2 est accessible, en préversion, par l'API Gemini ; le VLA et On-Device sont réservés aux partenaires en accès anticipé.

03 PROPRIÉTAIRE

### Figure · Helix 02 et Figure 03

**Helix 02 (27 janvier 2026)** pilote marche, accroupissement et manipulation dans un même modèle, avec des caméras dans les paumes et des capteurs tactiles de 3 g au bout des doigts. Figure a montré 61 actions enchaînées en 4 minutes autour d'un lave-vaisselle.Le modèle ne tourne que sur les robots de Figure.

» LES CERVEAUX OUVERTS, ET L'EUROPE

**Le code ouvert rattrape les modèles fermés d'hier.** LeRobot est devenu la boîte à outils commune ; les chiffres de performance ci-dessous sont ceux des éditeurs.

01 APACHE 2.0

### NVIDIA · GR00T N1.7

Accès anticipé le 17 avril 2026 puis disponibilité générale ; **usage commercial autorisé**. 3 Md de paramètres, pré-entraîné sur 20 854 h de vidéo humaine égocentrique. Validé sur Unitree G1, un robot bimanuel YAM et AgiBot Genie 1 ; export ONNX/TensorRT vers Jetson.

02 TOUT OUVERT

### Ai2 · MolmoAct 2

5 mai 2026 : poids, code et jeu de données bimanuel (plus de 720 h) publiés ; jusqu'à 37 fois plus rapide que MolmoAct. Ai2 annonce 87,1 % de réussite sur un Franka réel contre 45,2 % pour À0.5 (**chiffres auto-déclarés**).

03 OUVERT

### Hugging Face · SmolVLA et LeRobot

SmolVLA (450 M de paramètres, juin 2025) tourne sur un GPU grand public. LeRobot (27,6 k étoiles) fédère À0, À0.5, GR00T N1.7, SmolVLA et MolmoAct 2, avec des robots bon marché comme le bras SO-100, environ 100 $.

04 APACHE 2.0

### Physical Intelligence · openpi

Poids publics de À0, ~~FAST~~ et À0.5. À0.5 (22 avril 2025) généralise à des logements jamais vus pendant l'entraînement : environ 400 h de données collectées dans une centaine de logements.

05 FIGÉ

### OpenVLA · la référence historique

Modèle de 7 Md de paramètres (Llama 2 et Prismatic) entraîné sur Open X-Embodiment, qui a démocratisé la recherche en 2024. Son dépôt n'a plus évolué depuis le 23 mars 2025 : en pratique, openpi, GR00T, SmolVLA et MolmoAct 2 l'ont remplacé.

06 ÉCOSYSTÈME

### Europe · Hugging Face, Neura, Agile Robots

Hugging Face a racheté le français **Pollen Robotics** (14 avril 2025) et vend le petit robot Reachy Mini (399 ou 499 $). En Allemagne, **Neura Robotics** (humanoïde 4NE1) et **Agile Robots**, partenaire de Gemini Robotics 2, sont les principaux acteurs ; le français Enchanted Tools figurait parmi les premiers testeurs de Gemini Robotics.

RÉALITÉ / 03

# Démonstrations contre déploiements

Un seul déploiement industriel est solidement documenté. Le reste relève de livraisons chinoises, de précommandes, de calendriers qui glissent et de valorisations sans revenu publié.

**Le décalage reste net entre vidéos et robots livrés.** BMW est le déploiement industriel le mieux documenté ; partout ailleurs, les chiffres sont des annonces d'éditeurs ou de la presse spécialisée.

01 DÉPLOYÉ

### BMW et Figure : la preuve industrielle

Figure 02 a contribué à produire **plus de 30 000 BMW X3 en environ 10 mois**. Figure 03 et Helix 02 trient des pièces en logistique de séquencement à Spartanburg (annonce BMW du 25 juin 2026), sans nombre de robots communiqué.

02 LIVRÉ

### Unitree : des livraisons, une introduction en Bourse

Plus de 5 500 humanoïdes livrés en 2025 (32 % du marché selon son prospectus). Entrée à la Bourse de Shanghai le 19 août 2026 : cours multiplié par 5 à 7 le premier jour, selon les sources.

03 TÉLÉOPÉRÉ

### 1X NEO : précommandes et téléopération

Précommandes ouvertes le 28 octobre 2025 (20 000 $, ou 499 $ par mois). Les tâches que le robot ne maîtrise pas passent en « Expert Mode » : un téléopérateur humain prend la main, ce qui pose aussi des questions de vie privée. **Aucune livraison client vérifiée** à la mi-juillet 2026.

04 ANNONCÉ

### Tesla Optimus : toujours annoncé

Selon la presse spécialisée, ni l'Optimus V3 n'a été présenté ni la production lancée : le rapport trimestriel du 22 juillet 2026 dit « plus tard cette année ». Les calendriers ont déjà glissé plusieurs fois (sources secondaires).

05 VALORISATIONS

### L'argent : des valorisations sans revenu

Figure est valorisée 39 Md$ (septembre 2025). Physical Intelligence est passée de 2,4 Md$ (2024) à 5,6 Md$ (novembre 2025) et négociait en mars 2026 une levée d'environ 1 Md$ (clôture non confirmée). Neura Robotics annonce jusqu'à 1,4 Md$. **Aucune ne publie de revenu significatif lié aux VLA.**

06 RECHERCHE

### Maison et entrepôt : surtout des démonstrations

À0.5 nettoie cuisines et chambres dans des logements jamais vus, sur des tâches de 10 à 15 minutes ; À*0.6 prépare des espressos pendant 18 h. Ce sont des résultats d'éditeurs, obtenus sur des tâches choisies : ils ne disent pas encore la fiabilité d'un robot livré chez un client.

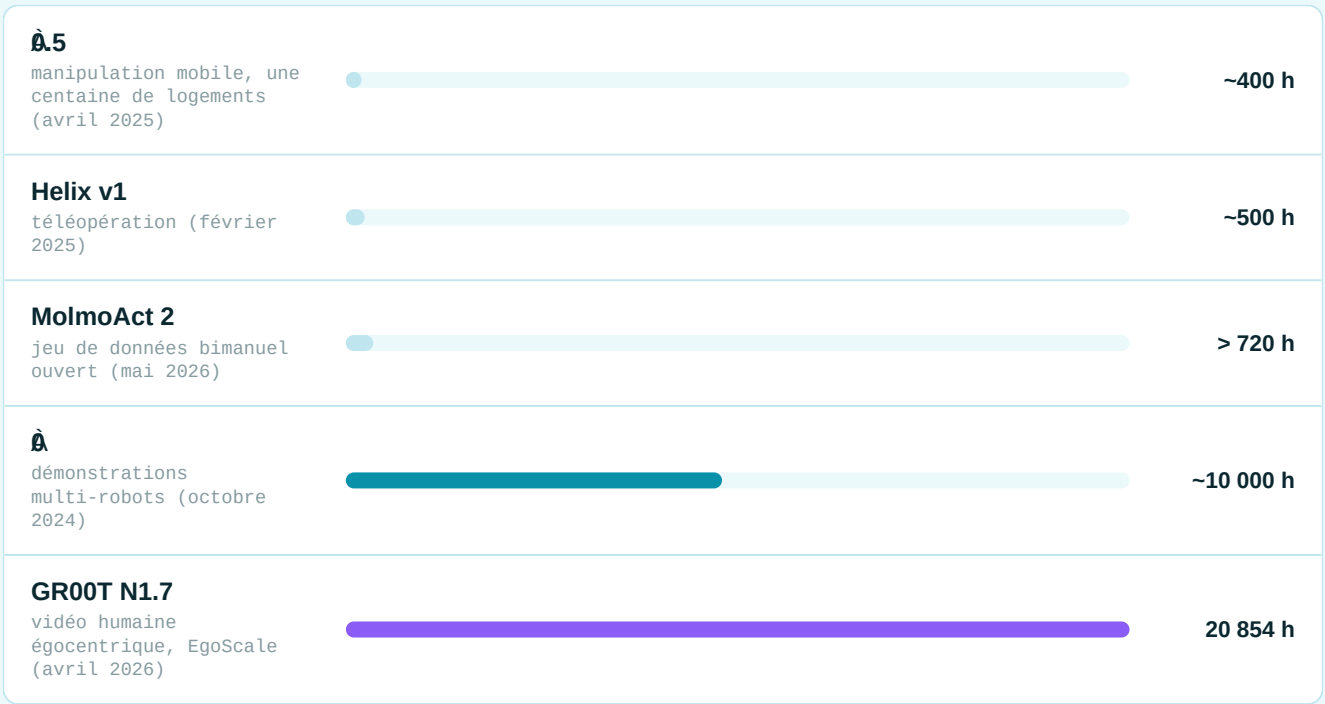
VERROUS / 04

# Les verrous : données, fiabilité, temps réel, sécurité

Le modèle n'est plus seul en cause. Ce qui manque désormais, c'est de quoi l'entraîner, la certitude qu'il tienne des heures, le calcul embarqué et un cadre de sécurité.

> **Le carburant change. La téléopération reste rare et chère ; la vidéo humaine égocentrique devient le levier principal ; la simulation complète : NVIDIA a généré 780 000 trajectoires synthétiques (environ 6 500 h) en 11 heures de calcul pour GR00T N1.**

» COMBIEN D'HEURES DE DONNÉES POUR ENTRAÎNER UN VLA ?




Volumes annoncés par les éditeurs, de natures différentes (téléopération, vidéo humaine, démonstrations) et donc non strictement comparables.

» QUATRE VERROUS

01 VERROU 1

### Les données

Collecter des démonstrations robotiques est lent, coûteux et risqué : aucune architecture ne compense un manque de données variées. D'où le pari sur la **vidéo humaine** (GR00T N1.7) et sur la simulation, qui reste imparfaite.

02 VERROU 2

### La fiabilité

Avec À*0.6le robot s'améliore sur ses propres essais : le débit a plus que doublé sur la préparation d'espresso. Deux briques s'ajoutent en mars 2026 : une mémoire à court et long terme (MEM) et du RL en ligne à partir d'un « RL token ». Hors distribution, la généralisation reste fragile.

03 VERROU 3

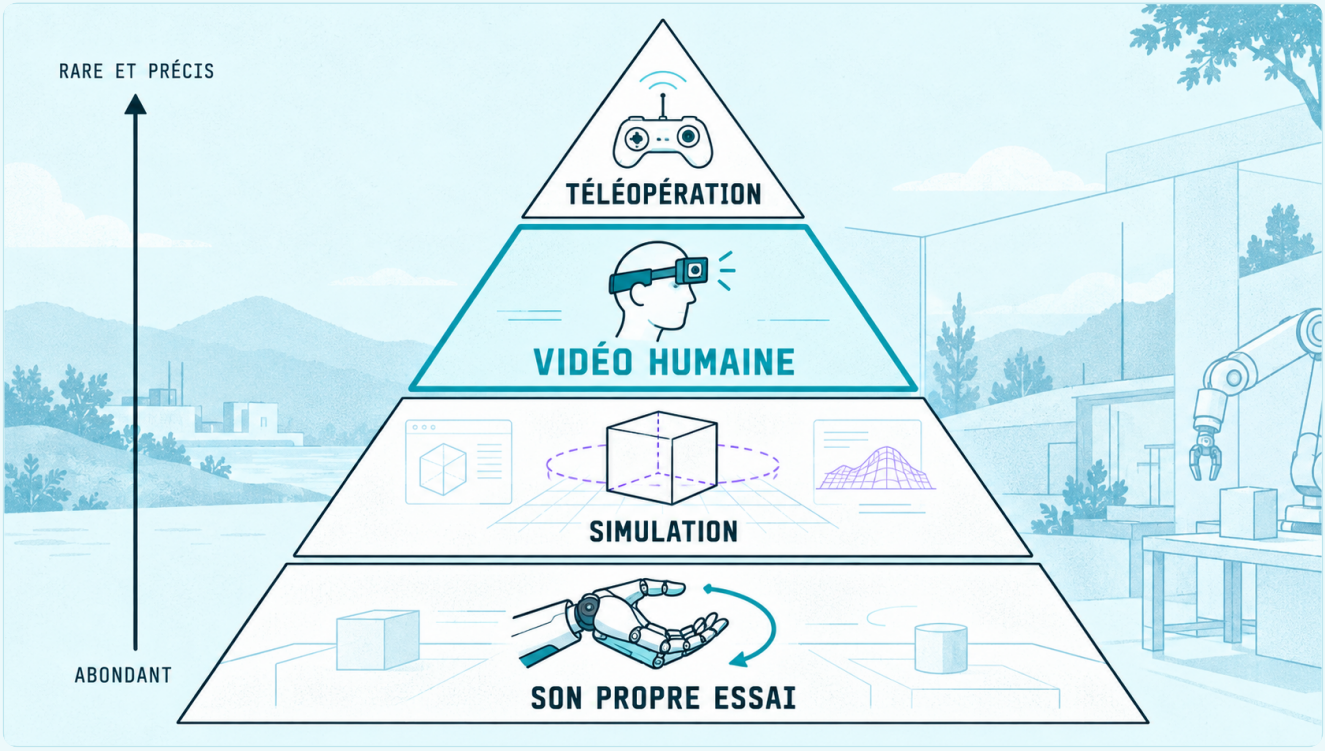
### Le temps réel et le matériel embarqué

Le robot est bien plus contraint en calcul et en énergie que le centre d'entraînement. **Jetson AGX Thor** (NVIDIA, 25 août 2025, dès 3 499 $) apporte 7,5 fois le calcul de Jetson Orin et rend plausible l'exécution locale de VLA de plusieurs milliards de paramètres ; Figure, Agility et Boston Dynamics l'ont adopté.

04 VERROU 4

### La sécurité et la réglementation

DeepMind publie ses propres tests (benchmark ASIMOV, puis ASIMOV-Agentic avec Gemini Robotics 2, détection de proximité humaine, arrêt automatique). En Europe, le **Règlement Machines (UE) 2023/1230** s'applique au 20 janvier 2027 ; l'Omnibus IA repousse les obligations « haut risque » de l'AI Act au 2 août 2028 pour les produits réglementés et exempte largement les machines déjà couvertes. Un ancien responsable sécurité de Figure a porté plainte en novembre 2025 au sujet de la force des robots.




Quatre sources de données pour un VLA : de la plus rare et précise à la plus abondante.

HORIZON / 05

# L'horizon : de l'acquis au spéculatif

Le chemin parcouru depuis RT-1, puis les échéances de fin 2026 à 2030, chacune classée par degré de certitude.

**Les jalons jusqu'à septembre 2026 sont des faits datés. Les échéances qui suivent sont des anticipations : leur degré de certitude est une appréciation éditoriale de ce rapport, sauf mention contraire. Toute date future est un objectif annoncé par un acteur, jamais une prévision du site.**

» FRISE : DE RT-1 À 2030, AVEC UN DEGRÉ DE CERTITUDE
Échelle : ACQUIS · PROBABLE · PLAUSIBLE · INCERTAIN · SPÉCULATIF (de moins en moins certain)

| | |
|---|---|
| 2022-2023 | **RT-1, RT-2, Open X-Embodiment**<br>RT-1 (décembre 2022), RT-2 (juillet 2023) qui donne naissance au terme VLA, puis 22 types de robots réunis dans Open X-Embodiment (octobre 2023). |
| 2024 | **OpenVLA et Ø**<br>OpenVLA 7B ouvert (juin) ; Ø introduit le flow matching (octobre). |
| Févr.-juin 2025 | **Helix, GR00T N1, Ø.5, SmolVLA, RTC, On-Device**<br>Figure, NVIDIA et Google DeepMind sortent leurs modèles ; Ø.5 généralise à des logements inconnus ; Hugging Face rachète Pollen Robotics. |
| Août-sept. 2025 | **MolmoAct, Jetson Thor, Gemini Robotics 1.5**<br>Un modèle ouvert qui raisonne en 3D, un calcul embarqué 7,5 fois plus puissant, un robot qui « réfléchit avant d'agir » ; Figure valorisée 39 Md$. |
| Nov. 2025 | **Ø.6 : le RL revient**<br>Le robot s'améliore sur ses propres essais. |
| Janv.-mai 2026 | **Helix 02, Ø.7, GR00T N1.7, MolmoAct 2**<br>Le corps entier, un modèle que l'on coache en langage naturel, un modèle ouvert entraîné sur 20 854 h de vidéo humaine, un modèle entièrement ouvert. |
| Juin-sept. 2026<br>AUJOURD'HUI | **Aujourd'hui**<br>Figure 03 chez BMW (juin), Gemini Robotics 2 (30 juillet), entrée en Bourse d'Unitree (19 août) : des cerveaux généralistes pour le corps entier, des déploiements comptés en dizaines ou centaines d'unités hors Chine. |
| **Fin 2026**<br>INCERTAIN | **Début de production d'Optimus, premières livraisons de NEO**<br>Objectifs annoncés par Tesla et 1X, dont les calendriers ont déjà glissé. |
| **2027**<br>PROBABLE | **Le Règlement Machines s'applique**<br>Le 20 janvier 2027, le règlement (UE) 2023/1230 devient applicable : le texte est en vigueur. |
| **2027**<br>INCERTAIN | **NEO hors des États-Unis, ligne Optimus au Texas**<br>Annonces de 1X et objectif de Musk, sans calendrier ferme. |
| **2028**<br>PROBABLE | **L'AI Act pour les produits réglementés**<br>L'Omnibus repousse les obligations « haut risque » au 2 août 2028 et exempte largement les machines déjà couvertes par le Règlement Machines. |
| **2030**<br>SPÉCULATIF | **Jusqu'à 10 000 NEO chez les clients d'EQT**<br>Accord 2026-2030 : une projection d'acteur, issue d'une source unique. |

**2030**
SPÉCULATIF

**Des robots domestiques autonomes, sans téléopération**
Le but affiché des acteurs du domestique : aucune source ne le date de façon crédible.

POUR ALLER PLUS LOIN **World Models : simuler un monde qui répond aux actions ›**

REPÈRES / 06

# Que faire de tout cela en 2026

Quatre réflexes devant une vidéo de robot spectaculaire.

» À RETENIR

### Quatre réflexes devant une vidéo de robot

01 Demander qui tient la télécommande : téléopération, « Expert Mode », séquence répétée ou autonomie réelle, et combien d'unités tournent chez un client.

02 Chercher les chiffres hors de l'éditeur : les comparaisons de performance publiées avec un nouveau modèle sont presque toujours auto-déclarées.

03 Pour prototyper, partir des modèles ouverts (GR00T N1.7 sous Apache 2.0, SmolVLA, MolmoAct 2) et de LeRobot sur du matériel bon marché, avant d'envisager un modèle fermé.

04 Ne pas bâtir un plan sur un calendrier d'annonce, et intégrer dès le départ le Règlement Machines (20 janvier 2027) dans toute étude de déploiement en Europe.

RÉFÉRENCES / 07

# Sources et ressources

[01] Physical Intelligence : À0.7a Steerable Model with Emergent Capabilities (16 avr. 2026)
[02] Physical Intelligence : À*0.6a VLA that Learns from Experience (17 nov. 2025)
[03] Physical Intelligence : À0.5a VLA with Open-World Generalization (22 avr. 2025)
[04] arXiv : À0a Vision-Language-Action Flow Model (2410.24164)
[05] Physical Intelligence : Real-Time Action Chunking (9 juin 2025)
[06] openpi : dépôt de Physical Intelligence (Apache 2.0)
[07] Google DeepMind : Gemini Robotics 2 (30 juil. 2026)
[08] Google DeepMind : Gemini Robotics 1.5 (25 sept. 2025)
[09] Google DeepMind : Gemini Robotics On-Device (24 juin 2025)
[10] Google AI for Developers : Robotics overview (Gemini Robotics-ER)
[11] Figure : Helix 02, full-body autonomy (27 janv. 2026)
[12] Figure : Helix, a VLA for Generalist Humanoid Control (20 févr. 2025)
[13] Figure : F.03 arrives at BMW (30 juin 2026)
[14] BMW Group : Figure 03 project in Spartanburg (25 juin 2026)
[15] Hugging Face / NVIDIA : Isaac GR00T N1.7 (17 avr. 2026)
[16] NVIDIA : Isaac-GR00T (dépôt, Apache 2.0)
[17] NVIDIA : Isaac GR00T N1, premier modèle ouvert pour humanoïdes (18 mars 2025)
[18] NVIDIA : Jetson AGX Thor disponible (25 août 2025)
[19] Ai2 : MolmoAct 2 (5 mai 2026)
[20] Ai2 : MolmoAct (12 août 2025)
[21] Hugging Face : SmolVLA (3 juin 2025)
[22] Hugging Face : rachat de Pollen Robotics (14 avr. 2025)
[23] Hugging Face : Reachy Mini
[24] Open X-Embodiment : jeu de données cross-embodiment (arXiv 2310.08864)
[25] Google DeepMind : RT-2, de la vision et du langage à l'action (juil. 2023)
[26] OpenVLA : dépôt du modèle ouvert 7B
[27] TechCrunch : À0.7et robots que l'on peut coacher (16 avr. 2026)
[28] TechCrunch : Physical Intelligence en discussion pour lever 1 Md$ (27 mars 2026)

[29] CNBC : Unitree bondit à son entrée en Bourse (19 août 2026)

[30] Electrek : Tesla repousse la présentation d'Optimus V3 (22 avr. 2026)

[31] The Robot Report : 1X lance les précommandes de NEO

[32] Humanoids Daily : 1X NEO, stratégie « human-in-the-loop »

[33] Gibson Dunn : EU AI Act Omnibus Agreement (calendrier et exemption des machines)

[34] Wikipedia : Vision–language–action model

[35] arXiv : VLA Models, Concepts, Progress, Applications and Challenges (2505.04769)

[36] arXiv : VLA Models for Robotics, a Review Towards Real-World Applications (2510.07077)