

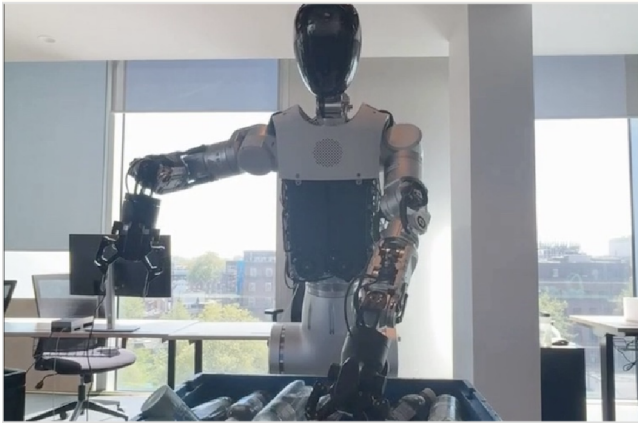
Pixels2Actions

MONDAY, JULY 6, 2026

PIXELS2ACTIONS · DAILY EDITION

33 STORIES

“Physical AI and robot learning, focused on how recorded video is used to train and supervise robot manipulation and control policies, reducing dependence on costly on-robot data collection. Learning policies from human a...”



LEAD STORY · THEROBOTREPORT · 18H AGO

Humanoid says KinetIQ Ascend reinforcement learning approaches human-level dexterity

Humanoid says its KinetIQ Ascend approach can reach 99.9% manipulation reliability at human speed and beyond for industrial tasks.

THEROBOTREPORT.COM

TOP NEWS

Michigan AI Data Center to Deploy Chinese Humanoid Robots to Replace Human Workers

A Michigan data center is preparing to deploy Chinese humanoid robots as companies continue pushing artificial intelligence (AI) from computer screens into the physical workplace. Hyperscale Data Inc. and its subsidiary, Omnipresent Robotics, are partnering...

CBNC · 17H AGO

Young lecturer brings AI, robotics into vocational education

For years, the 39-year-old teacher has worked to bring advanced technologies once found only in modern factories into vocational classrooms, helping bridge the gap between education and industry. Vietnam launches comprehensive reform to improve vocational...

VIETNAMPLUS · 1D AGO

China Deploys Humanoid Robots at Vietnam Border: Walker S2 Begins Live Customs Work

Humanoid robot border deployment crosses into live operations at China's Fangchenggang crossing, where UBTECH Walker S2 units handle cargo manifest scanning and traveler guidance under a 264-million-yuan contract. The split-brain AI architecture and Chinese...

TECHTIMES · 1D AGO

FROM THE NEWSLETTERS



The Queen's Journal – #37

Robotics updates bring several major launches from Automate 2026, with new developments in Physical AI, robot safety, industrial cobots, palletising, warehouse automation and fenceless robotics.

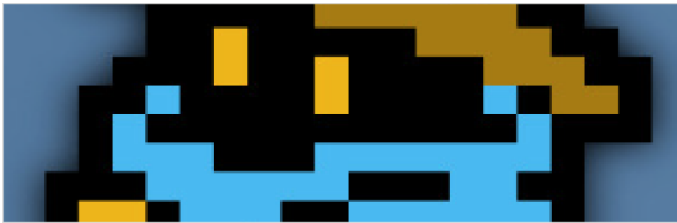
THE QUEEN'S JOURNAL · 4H AGO



EP.107 WEAVE'S ROBOT BRINGS AUTONOMOUS HOME CHORES TO MARKET

Appronik releases a new humanoid, Agility Robotics goes public as the first humanoid company & much more...

WE ALL ARE ROBOTS · ♥ 10 · 1 COMMENTS · 5H AGO



Robotics News & Literature (June 29th)

Apart of the series 'Another Fine Grasp'

PHYSICAL AI: ROBOTS, SCIENTIFIC COMPUTING, AND DIGITAL TWINS

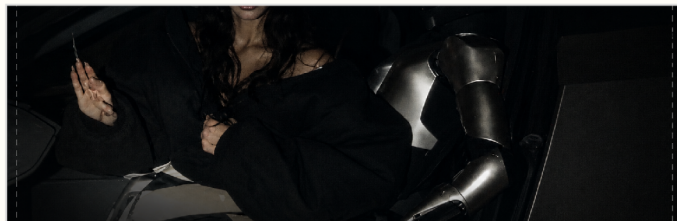
· ♥ 4



Robotics Pick & Shovel Part I: The Seven-Layer Enabling Stack

Robotics Pick & Shovel Part I of III

DARKHORSE'S RESEARCH · ♥ 4



Looks like everyone's getting turned on by robotics.

Hello to the uninvited.

♥ 16

NEW RESEARCH

ABot-Mo.5: Unified Mobility-and-Manipulation World Action Model

Mobile manipulation is a key capability for general-purpose robots, yet remains challenging for current embodied learning methods. VLA policies are typically reactive and lack explicit world modeling, while existing World Action Models (WAMs) are still poorly...

RONGHAN CHEN, YANDAN YANG, ZUOJIN TANG ET AL.

Local Motion Matters: A Deconstruct-Recompose Paradigm for Reinforcement Learning Pre-training from Videos

Pre-training on large-scale videos to improve reinforcement learning efficiency is promising yet remains challenging. Existing methods typically treat the agent as an indivisible entity, modeling motion patterns globally. Such global modeling is tightly...

JINWEN WANG, YOUFANG LIN, XIAOBO HU ET AL.

Unleashing More Actions via Action Compositional Training for VLA Models

Vision-Language-Action models excel at robotic manipulation, driven by the scale and diversity of demonstration data. However, standard training paradigms often cause VLA models to severely overfit to specific behavioral patterns, rendering them unable to...

KAI PENG, JIE LU, XIAOJIANG PENG

Structured 4D Latent Predictive Model for Robot Planning

Video predictive models are emerging as a powerful paradigm in robotics, offering a promising path toward task generalization, long-horizon planning, and flexible decision-making. However, prevailing approaches often operate on 2D video sequences, inherently...

ZHIYI LI, PEILIN WU, XIAOSHEN HAN ET AL.

Learning to Watch: Active Video Anomaly Understanding via Interleaved Policy Optimization

Video anomaly understanding (VAU) relies on sparse, context-dependent cues. However, existing passive paradigms suffer from observational aliasing, where static sampling fails to disambiguate semantically distinct events. To overcome this, we propose...

MENGJINGCHENG MO, JIAXU LENG, XINBO GAO

From World Models to World Action Models: A Concise Tutorial for Robotics

World models are increasingly used in embodied intelligence and generative simulation, yet their scope remains ambiguous across communities. This tutorial presents a design-space view of world models as action-conditioned predictive models that estimate the...

XIAOXIONG ZHANG, XIONG ZENG, WEI ZHANG

Human-Centric Transferable Tactile Pre-Training for Dexterous Robotic Manipulation

As an essential modality for dexterous and contact-rich tasks, tactile sensing provides precise force feedback that cannot be reliably inferred from vision. However, limited by hardware and data collection systems, existing datasets with tacity remain...

CHI ZHANG, PENGLIN CAI, ZIHENG XI ET AL.

Domain Arithmetic: One-Shot VLA Adaptation under Environmental Shifts

Vision-Language-Action (VLA) models often fail to perform the same learned tasks under environmental shifts, such as changes in camera pose and shifts to a different but similar robot (e.g., from Panda to UR5e). Adapting these models to the shifted...

TAEWOOK KANG, TAEHEON KIM, DONGHYUN SHIN ET AL.

THE DISCOURSE — X, REDDIT & HACKER NEWS

Forget lidar. One single camera. Runs in real time & is open source: A streaming 3D model that reconstructs...

@ILIRALIU_ · 2,550 LIKES · 284 REPOSTS

Today, we give robots a /skills library that self-evolves and compounds indefinitely! Introducing ASPIRE: a...

@DRJIMFAN · 769 LIKES · 109 REPOSTS

Talking with Gemma 4 31B!

Hi! I'm Andi from Hugging Face. This is a fully open-source and free to test/pull/modify demo I'm bringing today. It's a voice demo creating a pipeline of: \- Nvidia's parakeet \- Gemma 4 31B (served by cerebras!) \- My custom inference for Qwen3TTS It sees...

R/LOCALLLAMA · ▲ 612 · 109 COMMENTS

"What should I do?" - consider post-training

This is in response to the common post where OP has acquired some cool hardware and is wondering what to do with it. The standard response is always (1) download model X, (2) benchmark it on tps, (3) share screenshots. I argue this is boring and...

R/LOCALLLAMA · ▲ 604 · 128 COMMENTS

Palantir is a free org on HF with 0 open-source models and 0 public datasets shared

From clem 🐼 on X:

<https://x.com/ClementDelangue/status/2072683707001930215> From Palantir on X (video):

<https://x.com/PalantirTech/status/2072326189079757277> The information: Palantir CEO Says Some U.S. Government Customers Switched to Open Source AI...

R/LOCALLLAMA · ▲ 486 · 68 COMMENTS · 2D AGO

Oomwoo, an open-source robot vacuum you build yourself

450 POINTS · 88 COMMENTS

The consequences of "filters" in models (follow-up - KREA2)

(PROMPTS AT THE END) As a follow-up to a post I published yesterday, which seemed to be helpful to some people, I decided to make a more thorough comparison between the "pure" Krea2 Turbo model and the same model with three different de-censor LoRAs: two tiny...

R/STABLEDIFFUSION · ▲ 392 · 86 COMMENTS

Introducing LongCat-2.0 - , a large-scale MoE language model with 1.6 trillion total parameters and ~48 billion activated per token. This was the stealth model that was on Openrouter under the name 'owl-alpha'.

R/LOCALLLAMA · ▲ 373 · 66 COMMENTS

The real skill in AI video is picking the right reference TYPE per shot, not the model

After enough shots I stopped thinking about which model and started thinking about which reference type to feed it per shot. Same model, but the reference you hand it decides the shot, and matching the type to the shot is the actual skill. Three reference...

R/COMFYUI · ▲ 305 · 43 COMMENTS

UltraReal - LoRA for KREA2

This LoRA designed to reduce the typical *smooth/plastic AI look* and add more natural skin texture and realism to images. It works especially well for close-ups and medium shots where skin detail is important. It is trained on high-quality SFW and ****...

R/STABLEDIFFUSION · ▲ 303 · 138 COMMENTS

I benchmarked 13 models at 65K-128K context to find out what actually matters for agentic workloads

I benchmarked 13 models at 65K-128K context to find out what actually matters for agentic workloads — prefill dominates everything, and KV head count beats parameter count I've been running local LLMs for agentic workflows (tool use, coding agents, RAG) and...

R/LOCALLLAMA · ▲ 291 · 123 COMMENTS · 1D AGO

Krea 2 uses a terrible VAE, but luckily for us, PiD is here to save the day.

What is PiD -> <https://github.com/nv-tlabs/PiD> Krea 2 uses the worst VAE for DiT models (Qwen Image VAE), as a result, the final image is overly smooth, lacks detail and has poor color and contrast. The PiD method is designed to decode in pixel space...

R/STABLEDIFFUSION · ▲ 269 · 84 COMMENTS

As promised Krea 2 Turbo + "Raw" Quantized in FP8, MXFP8, NVFP4, INT8 and Convrot INT8!

Krea 2 Base & Turbo — Free Quantized Versions (FP8 / MXFP8 / NVFP4 / INT8 / ConvRot INT8) for All GPU Tiers Krea 2 just dropped and it's genuinely impressive — so I went ahead and quantized both variants for ComfyUI across every major format. All files...

R/STABLEDIFFUSION · ▲ 262 · 162 COMMENTS

Follow-up: DeepSeek V4 Flash on 2x RTX PRO 6000 finishes real coding tasks faster than Sonnet and Opus, at about Sonnet quality

This is a follow-up to post about which local models stay fast deep into long context and I learned a lot from people here. I kept measuring after that and it turned into a proper indie coding bench. With DeepSeek V4 Flash running on vLLM it lands around...

R/LOCALLLAMA · ▲ 246 · 94 COMMENTS · 2D AGO

Andrew Ng: "In 3-6 months, everyone will be using self-improving loops. No more prompting"

Andrew Ng recently said: "100% of my tasks are now done by AI agents. Hype has exceeded my expectations. Loops is next step. In 3-6 months, everyone will be using self-improving loops. No more prompting."* I think he's not too far off, you can already see...

R/ARTIFICIAL · ▲ 230 · 145 COMMENTS · 2D AGO

We chased a hallucinated quote through 30k training records, 4,600 transcripts, and our own system prompt. Turned out to be two separate bugs

Some of our customers noticed Inter-1 (our omni-modal social-signal model) would occasionally "hear" a quote that didn't exist. Feed it a video with zero audio and ask what was said, and it would sometimes report: "Yeah, Friday at five."* Verbatim. Same...

R/ARTIFICIAL · ▲ 214 · 45 COMMENTS