Rapport de veille hebdomadaire

Généré automatiquement le 12/04/2025

Geo4D: Leveraging Video Generators for Geometric 4D Scene Reconstruction

Date : 11/04/2025
Source : Feed for https://jamesg.blog/hf-papers.json

> Geo4D can be trained using only synthetic data while generalizing well to real data in a zero-shot manner. Geo4D predicts several complementary geometric modalities, namely point, depth, and ray maps. It uses a new multi-modal alignment algorithm to align and fuse thesemodalities, as well as multiple sliding windows, at inference time. It can be used to obtain robust and accurate 4D reconstruction of long videos.

Intelligence artificielle et services publics : la CNIL publie le bilan de son « bac à sable »

Date : 11/04/2025
Source : RSS - Actualités CNIL

> La CNIL publie les recommandations faites aux organismes accompagnés dans le cadre du 'bac à sable' 2023-2024. Le programme d'accompagnement personnalisé destiné aux acteurs pour déployer un projet innovant. Cette initiative s'inscrit dans la volonté de la CNIL de soutenir le développement d'une IA respectueuse des droits des personnes.

TAPNext: Tracking Any Point (TAP) as Next Token Prediction

Date : 11/04/2025
Source : Feed for https://jamesg.blog/hf-papers.json

> Tracking Any Point (TAP) in a video is a challenging computer vision problem. Existing methods for TAP rely heavily on complex inductive biases and heuristics, limiting their generality and potential for scaling. TAPNext casts TAP as sequential masked token decoding. Our model is causal, tracks in a purely online fashion, and removes tracking-specific biases. This enables TAP next to run with minimal latency and removes temporal windowing required by many existing trackers.

Scaling Laws for Native Multimodal Models Scaling Laws for Native Multimodal Models

Date : 11/04/2025
Source : Feed for https://jamesg.blog/hf-papers.json

> Building general-purpose models that can effectively perceive the world through multimodal signals has been a long-standing goal. Current approaches involve integrating separately pre-trained components, such as connecting vision encoders to LLMs. While such approaches exhibit remarkable sample efficiency, it remains an open question whether late-fusion architectures are inherently superior. In this work, we revisit the architectural design of native NMMs and conduct an extensive scaling study.

Compass Control: Multi Object Orientation Control for Text-to-Image Generation

Date : 11/04/2025

Source : Feed for https://jamesg.blog/hf-papers.json

> Existing approaches for controlling text-to-image diffusion models, while powerful, do not allow for explicit 3D object-centric control. In this work, we address the problem ofmulti-object orientation control. This enables the generation of diverse multi-object scenes with precise orientation control for each object. The model is trained on a syntheticdataset of procedurally generated scenes, each containing one or two 3D assets. However, direct training this framework results in poor orientation control as well as leads to entanglement among objects.

MonoPlace3D: Learning 3D-Aware Object Placement for 3D Monocular Detection

Date : 11/04/2025

Source : Feed for https://jamesg.blog/hf-papers.json

> MonoPlace3D is a novel system that considers the 3D scene content to create realistic augmentations. It learns a distribution over plausible 3Dbounding boxes. Subsequently, we render realistic objects and place them according to the locations s. The key obstacle lies in automatically determining realistic object placement parameters - including position, dimensions, and directional alignment when introducing synthetic objects into scenes. It's particularly difficult to generate realistic scene-aware augmented data for outdoor settings.

Kimi-VL Technical Report

Date : 11/04/2025

Source : Feed for https://jamesg.blog/hf-papers.json

> Kimi-VL is an efficient open-source Mixture-of-Experts (MoE) vision-language model. It offers advanced multimodal reasoning and strong agent capabilities. It effectively competes with cutting-edge efficient VLMs such as GPT-4o-mini, Qwen2.5-VL-7B, and Gemma-3-12B-IT.

Towards Visual Text Grounding of Multimodal Large Language Model

Date : 11/04/2025

Source : Feed for https://jamesg.blog/hf-papers.json

> TrIG is a novel task with a newly designed instruction dataset for benchmarking and improving the Text-Rich Image Grounding capabilities of MLLMs. We propose an OCR-LLM-human-interaction pipeline to create 800 manually annotated question-answer pairs as benchmark and a large-scale training set of 90$ synthetic data based on four diverse datasets. A comprehensive evaluation of various MLLM on our proposed benchmark exposes substauses.

MM-IFEngine: Towards Multimodal Instruction Following

Date : 11/04/2025

Source : Feed for https://jamesg.blog/hf-papers.json

> The Instruction Following (IF) ability measures how well Multi-modal Large Language Models (MLLMs) understand exactly what users are telling them. Existing multimodal instruction following benchmarks are simple with atomic instructions. The evaluation strategies are imprecise for tasks demanding exact output constraints. We present MM-IFEngine, an effective pipeline togenerate high-quality image-instruction pairs.

C3PO: Critical-Layer, Core-Expert, Collaborative Pathway Optimization for Test-Time Expert Re-Mixing

Date : 11/04/2025

Source : Feed for https://jamesg.blog/hf-papers.json

> Mixture-of-Experts (MoE) Large Language Models (LLMs) suffer from severely sub-optimal expert pathways. naive expert selection from pretraining leaves a surprising 10-20% accuracy gap forimprovement. We develop a novel class of test-time optimization methods to re-weight or "re-mixing" the experts in different layers jointly for each test sample. This leads to "Critical-Layer, Core-Expert,Collaborative Pathway Optimization (C3PO)".

DeepSeek-R1 Thoughtology: Let's about LLM Reasoning

Date : 11/04/2025

Source : Feed for https://jamesg.blog/hf-papers.json

> Large Reasoning Models like DeepSeek-R1 mark a fundamental shift in how LLMsapproach complex problems. Instead of directly producing an answer for a given input, DeepSeeks create detailed multi-step reasoning chains. This reasoning process is publicly available to the user, creating endless opportunities for studying the behaviour of the model and opening up the field of Thoughtology. Our findings paint a progressivelynuanced picture. Notably, we show DeepSeeker has a 'sweet spot' of reasoning, where extra inference time can impair model performance.

SoTA with Less: MCTS-Guided Sample Selection for Data-Efficient Visual Reasoning Self-Improvement

Date : 11/04/2025

Source : Feed for https://jamesg.blog/hf-papers.json

> This explicit step-by-step reasoning in MCTS enforces the model to think longer and better identifies samples that are genuinely challenging. We filter andretain 11k samplers and filter and retrain the model using the same set of samples. The main challenge remains in quantifying sample difficulty to enable effective data filtering. We propose a novel way of repurposing Monte Carlo Tree Search (MCTS) to achieve that.

VisualCloze: A Universal Image Generation Framework via Visual In-Context Learning

Date : 11/04/2025

Source : Feed for https://jamesg.blog/hf-papers.json

> The current mainstream approach remains focused on task-specific models, which have limited efficiency when supporting a wide range of different needs. Unlike existing methods that rely on language-based taskinstruction, we integratevisual in-context learning, allowing models to identify tasks from visualdemonstrations. Meanwhile, the inherent sparsity of visual task distributionshampers the learning of transferable knowledge. To tackle these challenges, we propose VisualCloze, a universal imagegeneration framework, which supports a range of in-domain tasks.

MOSAIC: Modeling Social AI for Content Dissemination and Regulation in Multi-Agent Simulations

Date : 11/04/2025

Source : Feed for https://jamesg.blog/hf-papers.json

> We present a novel, open-source social network simulation framework, MOSAIC, where generative language agents predict user behaviors such as liking, sharing, and flagging content. This simulation combines LLM agents with a directed social graph to analyze emergent deception behaviors and gain a better understanding of how users determine the veracity of online social content. Within this framework, we evaluate three different content moderation strategies and find that they not only mitigate the spread of untrue content but also increase user engagement.

[VCR-Bench: A Comprehensive Evaluation Framework for Video

Chain-of-Thought Reasoning](https://huggingface.co/papers/2504.07956) **Date** : 11/04/2025

Source : Feed for https://jamesg.blog/hf-papers.json

> VCR-bench is a benchmark designed to comprehensively evaluate LVLMs' Video-Chain-of-Thought Reasoning capabilities. VCR-Bench comprises 859 videos spanning a variety of video content and durations. Each pair is manually annotated with a stepwise CoT rationale, where every step is tagged to indicate its association with the perception or reasoning capabilities. The CoT score is used to assess the entire CoT process based on the stepwise tagged CoT rationals.

HoloPart: Generative 3D Part Amodal Segmentation

Date : 11/04/2025

Source : Feed for https://jamesg.blog/hf-papers.json

> 3D part amodal segmentation is a crucial task for 3D content creation and understanding. Existing 3D part segmentation methods only identify visible surface patches, limiting their utility. We introduce new benchmarks based on the ABO and PartObjaverse-based model. We leverage existing 3D segmentation to obtain initial, incomplete part segments. Second, we introduce a novel diffusion-based models to complete these segments into full 3D parts.

New Learning Pathway for Data Architects: Upskill on Data Platforms, AI and Governance

Date : 09/04/2025

Source : Databricks

> Databricks launches a new learning pathway to help data architects upskill in critical areas like modern data platforms, AI technologies, and governance. Early-bird savings of $600 — ends April 30. JUNE 9–12 | SAN FRANCISCO Our biggest Summit yet.Early-bird Savings of $6,000 — end April 30, early-bird discount of $1,000.

Lancement d’une concertation sur les outils d’enregistrement et de relecture de session de navigation

Date : 09/04/2025

Source : RSS - Actualités CNIL

> La CNIL organise une concertation avec les parties prenantes sur les outils d’enregistrement et de relecture de session de navigation. Ces outils, également appelés outils de rejeu de session, servent à reconstituer le parcours complet de navigation d’un utilisateur sur un site web ou uneapplication mobile.

Hugging Face and Cloudflare Partner to Make Real-Time Speech and Video Seamless with FastRTC

Date : 09/04/2025

Source : Hugging Face - Blog

> Hugging Face builtFastRTC to let AI developers build low-latency AI-powered audio and video streams with minimal Python code. WebRTC-powered applications often face deployment challenges due to the need for specialized TURN servers. As a preview of what you can build with FastRTC and Cloudflare, check out this voice chatappbuilt with Meta's new Llama 4 model.

The Power of Fine-Tuning on Your Data: Quick Fixing Bugs with LLMs via Never Ending Learning (NEL)

Date : 08/04/2025
Source : Databricks

> Our biggest Summit yet. Early-bird savings of $600 — ends April 30. JUNE 9–12 | SAN FRANCISCO Our biggest Summit Yet. early-bird Savings of $6,000 — ends May 1, 2015. inMosaic AI Research Summary:LLMs have revolutionized software development by increasing the productivity of programmers. Despite off-the-shelf LLMs being trained on a significant amount of code, they are not perfect. In this blog, we demonstrate how fine-tuning a small open-source LLM on interaction data enables state-of- the-art accuracy and low cost.

Databricks Wins 2025 Google Cloud Partner of the Year Award

Date : 08/04/2025
Source : Databricks

> Databricks named 2025 Google Cloud Data & Analytics Partner of the Year for Smart Analytics.Recognized for AI innovation, customer success, and deep product integration.Join us at Google Cloud Next and Data + AI Summit to learn what's next. Early-bird savings of $600 — ends April 30. JUNE 9–12 | SAN FRANCISCO Our biggest Summit yet.

Applications mobiles : la CNIL publie une version modifiée de ses recommandations pour mieux protéger la vie privée

Date : 08/04/2025
Source : RSS - Actualités CNIL

> La CNIL diffusait ses recommandations sur les applications mobiles, adoptées le 18 juillet 2024. Elle publie aujourd'hui une version mise à jour, après y avoir apporté des corrections mais sans en changer le fond. Le document adopté a fait l'objet de quelques modifications, non substantielles.

Arabic Leaderboards: Introducing Arabic Instruction Following, Updating AraGen, and More

Date : 08/04/2025
Source : Hugging Face - Blog

> Arabic-Leaderboards is a comprehensive and unified space for all Arabic evaluations and tasks. It is meant to serve as a central hub covering a broad spectrum of evaluations, for models across modalities. Currently, it has AraGen-03-25 and Arabic Instruction Following as live leaderboards. We plan to expand this space with more leaderbo with more model evaluations in the coming months.

Demandes d'autorisation en santé : bilan pour l'année 2024 de l'action de la CNIL

Date : 07/04/2025
Source : RSS - Actualités CNIL

> En 2024, la CNIL a reçu 619 demandes d'autorisation pour des traitements de données de santé, en hausse de 20% par rapport à 2023. La qualité des dossiers s'améliore, entraînant une réduction des délais d'instruction. La CNIL poursuivi son engagement pour accompagner la mise en conformité.

Databricks Bengaluru: Scaling Innovation and Building the Future of Data & AI

Date : 07/04/2025
Source : Databricks

> Databricks' R&D hub in Bengaluru is expanding. Discover how our engineering teams are helping to solve the world's toughest problems. We're hiring top engineers to build next-gen AI and data platforms. JUNE 9–12 | SAN FRANCISCO Our biggest Summit yet. Early-bird savings of $600 — ends April 30. Back to the page you came from.

Announcing the APJ Databricks Smart Business Insights Challenge: Empowering Data-Driven Decision Making with AI and BI

Date : 07/04/2025
Source : Databricks

> The Databricks Smart Business Insights Challenge is a virtual hackathon for data analysts, BI professionals, data engineers, and scientists. The challenge provides a unique opportunity to work together, apply AI-driven BI tools, and solve real-world business problems for a chance to win exciting prizes. JUNE 9–12 | SAN FRANCISCO Our biggest Summit yet. Early-bird savings of $600 — ends April 30.

Introducing Meta's Llama 4 on the Databricks Data Intelligence Platform

Date : 05/04/2025
Source : Databricks

> Llama 4 Maverick is now available on Databricks across AWS, Azure, and GCP. Develop and deploy fast, cost-effective, domain-specific agents, copilots, and RAG pipelines that use Mosaic AI. Built-in governance, including built-in logging, rate limiting, PII detection, and policy guardrails, helps to ensure safe use in production.
