

Act as a personal blog editor. I want you to review the grammar and expressions of the text. If necessary provide corrections and try to make the post easier to read, and appealing for a general and technical audience. Using more engaging language to capture the reader's interest. Also, consider breaking up longer sentences into shorter ones for easier reading. Do not expand the ideas of the text. If you have any suggestions write them after three dollar signs at the end of the text.

Text:

[GitHub - vikhyat/moondream: tiny vision language model](#)

[Fuyu-8B: A Multimodal Architecture for AI Agents](#)

[LLaVA-NeXT: Improved reasoning, OCR, and world knowledge | LLaVA](#)

<https://huggingface.co/blog/idefics2>

[01-ai/Yi-VL-6B · Hugging Face](#)

[GitHub - QwenLM/Qwen-VL: The official repo of Qwen-VL \(通义千问-VL\) chat & pretrained large vision language model proposed by Alibaba Cloud.](#)

[01-ai/Yi-VL-34B · Hugging Face](#)

[GitHub - THUDM/CogVLM: a state-of-the-art-level open visual language model | 多模态预训练模型](#)

[Kosmos 2 - a Hugging Face Space by_ydshieh](#)

[GitHub - ByungKwanLee/MoAI: Official PyTorch implementation code for realizing the technical part of Mixture of All Intelligence \(MoAI\) to improve performance of numerous zero-shot vision language tasks. \(Under Review\)](#)

For SD:

[Img2prompt | Methexis Inc | AI model details](#)

Clip interrogator follows an interesting design to create a bento service

[GitHub - pharmapsychotic/clip-interrogator: Image to prompt with BLIP and CLIP](#)

[GPT-4V Image Captioner. Installing and Running. | by Yushan777 | Medium](#)

Tuning

[https://colab.research.google.com/github/google-research/big_vision/blob/main/big_vision/configs/proj/paligemma/finetune_paligemma.ipynb](#)

[https://wandb.ai/byyoung3/ml-news/reports/How-to-Fine-Tune-LLava-on-a-Custom-Dataset--Vmlldzo2NjUwNTc1](#)