# Golia-v2

Versão atual: golia-2504

**Visão Geral**

Golia-v2 é a variante menor do modelo de visão e linguagem (VLM) na família Amazônia IA. Foi projetado para fornecer fortes capacidades multimodais de forma eficiente, visando configurações de hardware acessíveis enquanto mantém desempenho competitivo, especialmente em compreensão de documentos e perguntas e respostas visuais.

- Arquitetura do Modelo: Arquitetura similar ao Golia-v1/Amazônia IA (LLM + ViT), mas redimensionada para maior eficiência. Processa imagens em resoluções nativas.
- Características-chave e Capacidades Principais:
    - Oferece funcionalidades essenciais de visão e linguagem dimensionadas para seu tamanho: Compreensão Multimodal, Análise Visual e de Documentos Avançada, Interação Agêntica, Raciocínio e Saída Aprimorados.
    - Desempenho notavelmente forte para seu tamanho em compreensão de documentos, OCR e tarefas de perguntas e respostas visuais.

**Especificações Técnicas**

- Modalidades de Entrada: Texto, Imagem, Vídeo
- Modalidades de Saída: Texto
- Janela de Contexto: Suporta até 128K tokens. Máximo recomendado de 32K tokens para tarefas de visão e linguagem. Contate WideLabs se janelas de contexto maiores forem essenciais.
- Idiomas: Multilíngue (dezenas de idiomas). O desempenho em idiomas específicos pode ser aprimorado via fine-tuning.

**Desempenho**

Entrega resultados impressionantes para seu tamanho, oferecendo uma opção VLM poderosa mas acessível. Altamente competitivo, particularmente forte em benchmarks de documentos e perguntas e respostas visuais.

| **Benchmark Category** | **Benchmark** | **Score** | **Notes** |
|---|---|---|---|
| Vision (Doc/Chart) | ChartQA | 86.5 | Chart Visual Question Answering |
| Vision (Doc/Chart) | DocVQA | 92.1 | Document Visual Question Answering |
| Vision (Doc/Chart) | InfoVQA | 80.8 | Information Visual Question Answering |
| Vision (General) | MMMU | 57.9 | College-level Multimodal Problems |
| Vision (General) | MMMU-Pro | 37.4 | Harder College-level Multimodal Problems |
| Vision (Math) | MathVision | 24.9 | Visual Math Reasoning |
| Vision (Math) | MathVista | 66.7 | Visual Math Reasoning |
| Vision (OCR & Reasoning) | OCR-BenchV2 | 54.2 | Comprehensive OCR & Reasoning |

A complete and updated leaderboard comparing all Amazônia IA models will be available soon at: [https://huggingface.co/spaces/widelabs/leaderboard](https://huggingface.co/spaces/widelabs/leaderboard)

**Implantação**

- Framework: vLLM
- Hardware (BF16): 1x A10
- Integração de API: API compatível com OpenAI
- Formatação de Entrada: Lista de dicionários (`role`, `content`). O conteúdo do usuário é uma lista de dicionários (`type`: ”text”/”image”/”video”, `data`: string/URL/caminho/base64).
- Engenharia de Prompt: Princípios padrão. Aproveite detalhes/formatação aprimorados. Combine instruções claras de texto com contexto visual.
- Quantização: Geralmente não recomendada devido ao tamanho menor do modelo; a degradação de desempenho pode ser mais significativa. Contate WideLabs se a quantização for necessária.
- Fine-tuning: Suportado e pode alcançar desempenho notável em tarefas específicas. Contate WideLabs para serviços de fine-tuning.

- Limitações: Pode encontrar limites de desempenho em GPUs de menor capacidade, especialmente ao processar entradas de alta resolução ou contextos muito longos simultaneamente.