

PR #38714 完整报告

vllm-project/vllm

Add ibm-granite/granite-vision-3.3-2b to supported models documentation

合并时间: 2026-04-01 23:22

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/38714>

执行摘要

本 PR 在 `docs/models/supported_models.md` 文档中添加了 Granite Vision 模型条目，解决了用户因文档缺失而误认为该模型不受支持的问题。这是一个纯文档变更，风险极低，仅影响文档准确性，对代码库无直接影响。

功能与动机

根据 PR body，用户搜索文档中的“granite vision”时找不到任何信息，从而错误地得出结论认为该模型不受支持。实际上，Granite Vision 已通过 `LlavaNextForConditionalGeneration` 后端在 vLLM 中正常工作，并且已在 `tests/models/registry.py` 中注册，但完全缺失于 `supported_models.md` 文档。因此需要添加文档条目以提供准确的用户指引。

实现拆解

仅修改了一个文件 `docs/models/supported_models.md`。具体变更如下：

文件	变更内容
<code>docs/models/supported_models.md</code>	在 <code>LlavaNextForConditionalGeneration</code> 行中，将描述从“LLaVA-NeXT”更新为“LLaVA-NeXT, Granite Vision”，并在示例模型列表中添加 <code>ibm-granite/granite-vision-3.3-2b</code> 。

变更前后对比（基于 `patch_excerpt`）：
- | `LlavaNextForConditionalGeneration` | LLaVA-NeXT | T + I^{E+} | `llava-hf/llava-v1.6-mistral-7b-hf`,
`llava-hf/llava-v1.6-vicuna-7b-hf`, etc. | | | + |
- | `LlavaNextForConditionalGeneration` | LLaVA-NeXT, Granite Vision | T + I^{E+} | `llava-hf/llava-v1.6-mistral-7b-hf`, `llava-hf/llava-v1.6-vicuna-7b-hf`,
`ibm-granite/granite-vision-3.3-2b`, etc. | | | |

评论区精华

review 中几乎没有技术讨论：

- `gemini-code-assist[bot]` 评论：“This pull request updates the documentation for supported models by adding Granite Vision to the `LlavaNextForConditionalGeneration`

entry and including ibm-granite/granite-vision-3.3-2b as an example. I have nofeedback to provide.”

- mgoin 直接批准，无额外评论。

没有争议点或未解决疑虑，变更被顺利接受。

风险与影响

风险分析：风险极低。这是纯文档变更，不涉及代码逻辑、性能、安全或兼容性。唯一潜在风险是文档准确性，但 PR 说明模型已通过现有后端支持且已在测试注册表中，因此风险可忽略。

影响分析：影响范围仅限于文档用户。解决了用户因文档缺失而产生的困惑，提升了文档的完整性和准确性。对系统、团队或代码库无直接影响。

关联脉络

与近期历史 PR 的关联：

- PR #38722 和 #38723 同为文档类小修复，涉及 documentation 和 cleanup 标签，体现了团队对文档细节的持续维护。
- 本 PR 延续了多模态模型（如 Granite Vision）文档支持的完善趋势，与仓库中 multi-modality 标签的 PR 相呼应。

整体来看，这是一个简单的文档补充，旨在确保用户能准确获取模型支持信息，避免因文档缺失导致的误解。