LFM3-VL 2B Heretic: MLX 4-bit export

Lucas van de weerd | Lucasgitguy-co | 7 November 2026 | Version 1.0

Repository: Lucasgitguy-co/lfm3-vl-2b-heretic-mlx-4bit

Reviewed revision: d1121f749f6a5cdd004a8be2a4ba5dd5b37a5333

Abstract

This report documents the mlx 4-bit export of LFM3-VL 2B Heretic, its published lineage and deployment contract. It distinguishes format conversion from model training and preserves the limits of the available evaluation evidence.

1. Scope and related work

This technical report reconstructs the release from pinned repository artifacts and the publisher's earlier portfolio report. It is an AI-assisted documentation synthesis, not an independent rerun or a peer-reviewed publication. Reported measurements belong to the original release records. [1, 2, 3, 4]

The earlier portfolio report provides the organizing distinction between training, model modification, conversion and packaging. This report applies that distinction to one release and uses current records where later experiments supersede the historical overview.

2. Model and method

The export uses 4-bit affine quantization with group size 64 and an MLX-compatible loader. The card lists a 1.352 GB weight file and identifies the publisher's BF16 Heretic checkpoint as its parent. Quantization changes numerical representation and runtime requirements; this record does not document a separate training run. [1, 2, 3, 4]

3. Evidence and results

No task benchmark, before/after modification comparison, or independently reproduced inference result is documented for this specific artifact in the inspected release. File existence and configuration declarations are the available evidence; they do not establish behavioral improvement. [1, 2, 3, 4]

4. Limitations and intended use

The historical Heretic modification command, search configuration, data and before/after task measurements remain incomplete. The name does not establish measured refusal changes or a quality ranking. LFM Open License v1.0 terms continue to apply to the weights and derivatives; documentation licensing does not replace them. For image input, use the matching processor/projector and a runtime supporting this architecture.

5. Reproducibility and conclusion

The source snapshot above fixes what was reviewed. The linked source files and source-manifest.json record the evidence used, including file hashes. Consult the repository's usage instructions and recorded dependencies for execution. No new model training, inference benchmark, or claimed state-of-the-art result was introduced in preparing this report. Evidence status: artifact documentation.

References

[1] Lucasgitguy-co. README.md at reviewed revision.

[2] Lucasgitguy-co. docs/PROVENANCE.md at reviewed revision.

[3] Lucasgitguy-co. docs/USAGE.md at reviewed revision.

[4] Lucasgitguy-co. config.json at reviewed revision.

[5] Lucasgitguy-co. Local Model Adaptation Across Vision Diagrams Geometry and Game Decisions. Historical portfolio report, 5 November 2026; its proposed v2 experiment is superseded by later release records.