NAME

llava-quantize — CLIP model quantizer

SYNOPSIS

llava-quantize [options...] *INPUT OUTPUT FORMAT*

DESCRIPTION

llava-quantize makes LLaVA mmproj files smaller.

ARGUMENTS

The following positional arguments are accepted:

INPUT   Is the input file, which should be a CLIP model in the GGUF format using float16 values.

OUTPUT   Is the output file, which will be a CLIP model in the GGUF format using the desired number type.

FORMAT   Is the desired quantization format, which may be the integer id of a supported quantization type. See the quantization types section below for acceptable formats.

OPTIONS

The following options are accepted:

-h, **--help**
> Show help message and exit.

--version
> Print llamafile version.

QUANTIZATION TYPES

The following quantization types are available:

- 2 is Q4_0
- 3 is Q4_1
- 6 is Q5_0
- 7 is Q5_1
- 8 is Q8_0

SEE ALSO

llamafile(1)