

Si queremos montar un servidor de inteligencia artificial donde corramos Ollama + Open WebUI y basemos todo el procesamiento en un procesador (no en una tarjeta gráfica), estos son los requerimientos de memoria RAM necesarios para cada modelo LLM (sólo he puesto los más interesantes):

- **llama3 8b**: x GB de RAM.
- **llama3 8b-instruct-fp16**: x GB de RAM.
- **llama3 70b**: x GB de RAM.
- **llama3 70b-instruct-fp16**: x GB de RAM.
- **mistral 7b**: x GB de RAM.
- **mistral 7b-instruct-fp16**: x GB de RAM.
- **phi3 3.8b**: x GB de RAM.
- **phi3 3.8b-mini-128k-instruct-f16**: x GB de RAM.
- **phi3 14b**: x GB de RAM.
- **phi3 14b-medium-128k-instruct-f16**: x GB de RAM.