MAKO

Fast AI. Any Hardware.
Make Every GPU Count.

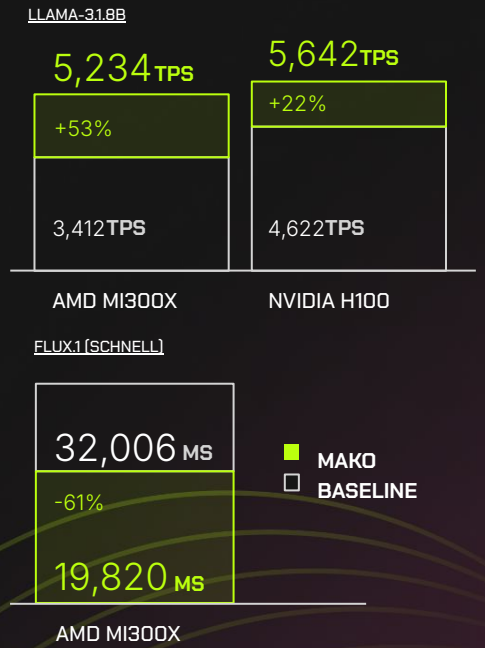
SEAMLESS INTEGRATION
Works out-of-the-box with any PyTorch or Hugging Face model—no code changes.

AUTOMATIC OPTIMIZATION
Auto-tunes GPU kernels & inference engines automatically with just one line of code.

CONTINUOUS IMPROVEMENT
Keeps learning and optimizing over time to make your AI faster & more efficient, 24/7.

SMART BACKEND SELECTION
Auto-selects best execution backend (vLLM, SGLang, TRT-LLM....)—no need for manual testing.

Proven Performance Gains:
Any Model, Hardware.



Built by Experts.
Backed by Leaders.

Waleed Atallah
CHIEF EXECUTIVE OFFICER

Łukasz Dudziak
CHIEF TECHNOLOGY OFFICER

Mohamed Abdelfattah, Ph.D
CHIEF SCIENCE OFFICER

Jeff Dean (CHIEF SCIENTIST - GOOGLE)
ADVISOR

Micah Villmow (FMR. PRINCIPAL SWE - NVIDIA, TENSORRT)
ADVISOR

M13 FLYBRIDGE
Neo
KS GLOBAL
Irregular expressions