Llama 3.1 8B Tokens per kJ <sub>→</sub>25000 e AMD MI300X Intel XPU **NVIDIA A100 NVIDIA GH200 NVIDIA H100**