



LLM recommended config matrix

HP ZBook Ultra G1a



Local Text-Based LLMs			Recommended ZBook Ultra configuration				Intended Use Case
AI Company	Model	Size	Processor & Graphics SOC	Memory	Storage	Default VRAM allocated to GPU	
Meta	Llama 3.3	70B	AMD Ryzen AI Max+ Pro 395 (16 Core CPU, 40 Core GPU)	128GB	4TB	75%	Evaluation, RAG Development, Inference
	Llama 3.2	11B	AMD Ryzen AI Max+ Pro 390 (12 Core CPU, 32 Core GPU) or greater	64GB or greater	2TB or greater	50% or greater	RAG Development, Prototyping, Inference
	Llama 3.2	3B	AMD Ryzen AI Max+ Pro 385 (8 Core CPU, 32 Core GPU) or greater	64GB or greater	2TB or greater	50% or greater	Prototyping, Finetuning, Light Development
	Llama 3.2	1B	AMD Ryzen AI Max+ Pro 385 (8 Core CPU, 32 Core GPU) or greater	32GB or greater	1TB or greater	50% or greater	Light Development, Inference
	Llama 3.1	8B	AMD Ryzen AI Max+ Pro 390 (12 Core CPU, 32 Core GPU) or greater	64GB or greater	2TB or greater	50% or greater	Inference, Prototyping
Alibaba Cloud	Qwen 3.0	32B, 72B	AMD Ryzen AI Max+ Pro 395 (16 Core CPU, 40 Core GPU)	128GB	4TB	50% or greater	Evaluation, Multilingual, RAG Development
	Qwen 3.0	7B, 14B	AMD Ryzen AI Max+ Pro 390 (12 Core CPU, 32 Core GPU) or greater	64GB or greater	2TB or greater	50% or greater	Multilingual, Inference, Prototyping
	Qwen 3.0	0.5B, 1.5B, 3B	AMD Ryzen AI Max+ Pro 385 (8 Core CPU, 32 Core GPU) or greater	64GB or greater	2TB or greater	50% or greater	Light Development, Finetuning
Google	Gemma 3	27B	AMD Ryzen AI Max+ Pro 390 (12 Core CPU, 32 Core GPU) or greater	64GB or greater	2TB or greater	50% or greater	Evaluation, RAG Development, Inference
	Gemma 3	12B	AMD Ryzen AI Max+ Pro 390 (12 Core CPU, 32 Core GPU) or greater	64GB or greater	2TB or greater	50% or greater	Inference, Prototyping
	Gemma 3	4B	AMD Ryzen AI Max+ Pro 385 (8 Core CPU, 32 Core GPU) or greater	64GB or greater	1TB or greater	50% or greater	Prototyping, Finetuning, Light Development

NOTE: 64GB VRAM and higher is highly recommended.