Generative Artificial Intelligence Training Equipment

GenWorks

Operating Program Specification

List		Specifications
OS	Kernel	Linux Kernel 6.x
	Desktop	Window Manager: OpenBox | Display Manager: LightDM | Panel: Tint2 System Monitor: conky | Bluetooth Manager: blueman Network Manager: network-manager | Graphics: NVIDIA Driver, CUDA V13.x, Runtime (cudart)
	CLI	Shell: Zsh, Oh-My-Zsh with powerlevel9k thema and nerd fonts Terminal Multiplexer: Tmux | Base Tools: fzf, bat, lsd
	Tool Chain	Python3, NodeJS, Java, Clang, GCC, LLVM
	IDE	Visual Studio Code for LLM
	Library	Data analysis: numpy, matplotlib, Pandas, SciPy, seaborn, plotly, Folium Image Processing: OpenCV, Mahotas, SimpleCV, Pillow, Scikit-Image, SimplelTK, Pgmagick Audio Processing: Pydub, Librosa, Soundfile, Pedalboard, Supriya, Libsox, Pyo, Pippi
AI	CUDA Toolkit, cuDNN, TensorRT	Pre-installed CUDA Toolkit, cuDNN, and TensorRT stack optimized TensorRT-LLM components for high-throughput FP8/INT8/FP16 large language model inference and KV-cache optimization
	Framework (PyTorch, TensorFlow, Keras, Scikit-Learn)	PyTorch, TensorFlow, Keras, and Scikit-Learn with full GPU support enabled Hugging Face Transformers and Diffusers libraries for large language models and diffusion-based image generation LangChain and related libraries for retrieval-augmented generation (RAG) and agent orchestration NVIDIA NeMo framework and NVIDIA TAO Toolkit containers for NVIDIA-optimized model training, adaptation, and deployment
	Algorithm	Machine Learning: Linear Regression, Logistic Regression, Decision Tree, SVM, Naive Bayes, KNN, K-means, Random Forest, Dimensionality Reduction (e.g., PCA), Gradient Boosting (XGBoost/LightGBM) and AdaBoost Object Detection: HOG, R-CNN and Fast/Faster R-CNN, SSD, YOLO family, RetinaNet Generative AI: Transformer encoder–decoder and decoder-only architectures, BERT, GPT-style language models, Llama 2/3 family, text embedding models, RAG pipelines, LoRA/QLoRA configuration utilities
	NVIDIA Model	NVIDIA TAO pre-trained computer vision models: TrafficCamNet, DashCamNet, PeopleNet, PeopleSemSegNet, VehicleMakeNet, VehicleTypeNet, LPDNet and related variants NVIDIA face and gaze models: FaceDetect, FaceDetectIR, FaceNet, FPENet, GazeNet, EmotionNet, HeartRateNet NVIDIA speech and audio models: WaveGlow, Flowtron, and fine-tuning configurations for Flowtron and other NeMo-based TTS models NVIDIA large language and generative models optimized for RTX 5090, including the Nemotron family and NVIDIA-optimized open LLM checkpoints Language identification models including LangID PearlNet and related NVIDIA NeMo language ID models

Component

• GenWorks

• Text Book

Artificial Intelligence Workstation with CARLA Simulator

Hardware Specification

List		Specifications
Body	Size	505 x 230 x 480 mm
	Weight	20kg
	Material	SGCC Steel, Ultra-clear Tempered Glass
	Color	White
Technical	CPU	Performance cores 8 + Efficient cores 16 Total Threads 32 MAX Turbo Frequency 6.00GHz Performance-core MAX Turbo Frequency 5.60GHz Efficient-core MAX Turbo Frequency 4.40GHz Performance-core MAX Base Frequency 3.20GHz Efficient-core MAX Base Frequency 2.40GHz Cache 36MB, Total L2 Cache 32MB
	GPU	CUDA Cores: 21,760 Boost Clock: 2.41GHz Base Clock: 2.01GHz Memory Size: 32GB (GDDR7) Memory Interface Width: 512bit Memory Bandwidth: 1,792 GB/s Ray Tracing Cores: 4th Generation (RTX Blackwell) Tensor Cores: 5th Generation, up to 3,352 AI TOPS NVIDIA Architecture: Blackwell NVIDIA Encoder (NVENC): 3× 9th Generation NVIDIA Decoder (NVDEC): 2× 6th Generation PCI Express Interface: PCIe 5.0 x16 Display Outputs (FE): 3× DisplayPort 2.1b, 1× HDMI 2.1b Maximum Resolution & Refresh Rate: 4K at 480Hz or 8K at 120Hz (with DSC, HDR) Maximum GPU Temperature: 90°C Total Graphics Power (TGP): 575W
	Base Board	Memory: 64GB DDR 5x, Tested Speed 6400MT/s Storage: PCIe 4.0 NVMe M.2 SSD 2TB. Reading MAX 7,000MB/s, Writing: MAX 6,500MB/s Audio: ALC1220 Codec, 32-bit/92kHz DAC, 120dB SNR DAC with Differential Amplifier Connectivity: RTL8125BG 2.5GbE LAN, Wi-Fi 6E AX211, Bluetooth v5.2 VRM: 16(VCORE) + 1(VCCGT) + 2(VCCIN_AUX) Interface: PCIe x16 x 3ea, PCIex1 M.2Socket(Key E) x 1ea USB 2.0 x 2ea, USB 3.2 (20Gbps) Type-C x 1ea, USB 3.2 (10Gbps) x 1ea, USB 3.2 (5Gbps) x 1ea,
	Cooling System	Water Block: Copper Housing, 75 x 56 mm ┃ Pump: Motor Speed 800~2800RPM Radiator: Aluminum, 121 x 394 x 27 mm Tube: Ultra-low Evaporation Rubber with Nylon Braided Sleeve, 400mm FAN: Top 120 x 120 x 26mm x3ea ┃ Front 120 x 120 x 26mm x 3ea ┃ Rear 140 x 140 x 26mm x 1ea
	Power	Cable Type : Full Modular DC Output : 1000W(+3.3V 20A, +5V 20A, +12V 83A, -12V 0.3A, +5Vsb 3A)
Display	Resolution	3840x2160 (4K UHD)
	Refresh Rate	60Hz
	HDR	HDR10 support
	Response speed:	4ms (GTG)
	Brightness	300cd
	Contrast ratio	5000:1
	I/O Port	HDMI x 2ea, USB 2.0 x 2ea, USB type-C x 1ea

Generative Artificial Intelligence Training Equipment

GenWorks

Pre-installed AI frameworks, including CUDA

AI Performance (INT8) 3,352 AI TOPS

- Built-in GPU based on Blackwell architecture with up to 24 cores and 32 threads CPU and 3,352 AI TOPS class
- Equipped with 64GB of main memory and 2TB NVMe SSD for processing large amount of data and model parameter
- High-efficiency cooling and power design considering GPU and CPU heating for stable performance of long-term full-load AI workload
- Provide minimal system package for the educational field and Ubuntu LTS distribution eliminating unnecessary services and reconfiguring generative AI development environment
- Support system image-based file system recovery mode and one-click initialization for returning to the initial state at any time after practice
- Support visual confirmation of generative AI results such as prompt-response process, image/video creation result, simulator screen through large 43-inch 4K monitor
- Pre-configured library and running environment specialized for generative AI, including Hugging Face Transformers/ Diffusers as well as CUDA-based PyTorch, TensorFlow
- Pre-loaded with open source-based language model and diffusion model and example project for representative generative AI practices such as text generation, code generation, image generation
- For professional application development, support extended features for Visual Studio Code-based integrated development environment and generative AI development, Jupyter, Git, Docker
- Provide generative AI educational content that can be applied directly to lecture to support consistent curriculum that connects theory, practice, project

Training Contents

- Overview of Generative Artificial Intelligence and Building Transformer, Practice Environment
- Deep Learning Model Basics (Forward/Backward, LeNet, Autograd)
- First Step to Computer in Understanding Natural Language. All About Tokenization and Attention
- Structure and Application of Generative Language Model Centered on GPT
- Make large-scale language model faster and lighter
- Learning GPT
- Understanding RAG_251202
- LLM, MCP Agent as Active Agent
- Understanding Multimodal LLM
- Principle and Practice of Diffusion
- Understanding Stable Diffusion

Subject to Change