

Ubiquitous AI. Powering Ingenuity.

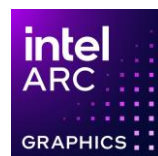
Features

ASUS UGen Intel® Arc™ Pro B70 delivers 32GB VRAM capacity, enabling efficient deployment of LLMs, RAG, AI Agents, image generation and enterprise AI applications without relying on cloud infrastructure.

- **GPU Peak TOPS (INT8):** 367 TOPS
- **GPU TFLOPS (FP32):** 22.94 TFLOPS
- **Xe-Cores:** 32
- **XMN Engines:** 256
- **Ray Tracing(RT) Units:** 32
- **PCIe® Support:** Gen 5 x 16 Native
- **Boost Clock:** up to 2800MHz
- **Memory:** 32GB GDDR6 with (ECC)
- **Memory Bandwidth:** 608 GB/s
- **Memory Interface:** 256-bit
- **Multi-GPU Scalability:** up to 128 GB VRAM
- **Display:** 3x DisplayPort 2.1
1x HDMI
- **Max. Resolution:** 7680 x 4320 @ 120 Hz
- **Power Cable:** 12V 2x6
- **TBP:** 290 W



intel ARC
PRO B70



DUAL BALL FAN BEARINGS

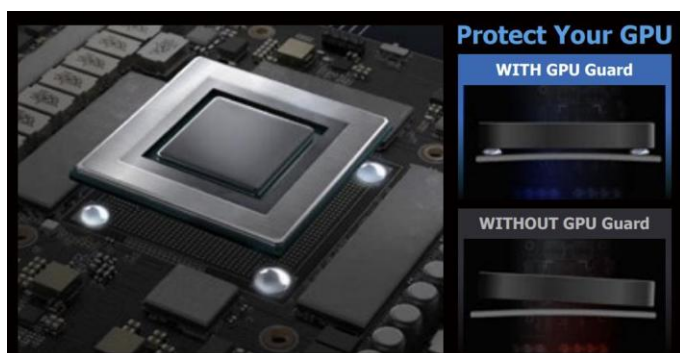
Dual ball bearings can last up to twice as long as sleeve-bearing designs.



*Image for reference only.

GPU GUARD

ASUS GPU Guard applies adhesive to secure all four corners to reduce the risk of cracks.



*Image for reference only.



UGen Intel® Arc™ Pro B70 32GB
UGEN-ARC-PRO-B70-32G

Ubiquitous AI. Powering Ingenuity.

Key Features

1. 32GB High-Capacity VRAM

Support larger AI models and longer context windows without excessive memory offloading.

2. Optimized for Local AI Inference

Purpose-built for LLM, RAG, AI Agents and Generative AI workloads.

3. Agentic AI Ready

Run autonomous AI agents locally with enhanced privacy and reduced operating costs.

4. Multi-GPU Scalability

Scale from 1, 2, or 4 GPUs to support larger models and higher user concurrency.

5. Enterprise Private AI Platform

Keep sensitive data on-premises while deploying AI services across departments.

6. RAG Knowledge Base Acceleration

Enable fast retrieval and generation from enterprise documents and internal knowledge repositories.

7. Generative AI Content Creation

Accelerate image generation, AI-assisted media production, and creative workflows.

8. Open Ecosystem Support

Compatible with OpenVINO™, Ollama, vLLM, llama.cpp, PyTorch, LangChain and popular AI frameworks.

9. Cost-Optimized AI Deployment

Deliver enterprise AI capabilities with a more accessible cost structure compared with traditional AI servers.

10. Concurrent Multi-User AI Serving

Support multiple users running AI assistants, chatbots and inference services simultaneously.

11. Flexible Deployment Options

Suitable for personal AI workstations, departmental AI servers and enterprise AI clusters.

12. Future-Ready AI Infrastructure

Designed to support next-generation LLMs, Reasoning Models and Multi-Agent AI systems.

Ubiquitous AI. Powering Ingenuity.

Supported Model Families

Built to accelerate today's most popular open-source AI models through Intel® vLLM and OpenVINO™ optimized ecosystem. Support coverage includes Large Language Models (LLMs), Vision Language Models (VLMs), Speech AI, OCR, Embedding Models and Agentic AI frameworks. [\[hub.docker.com\]](https://hub.docker.com)

Large Language Models (LLMs)

- DeepSeek™ Series
- Qwen™ Series
- Llama™ Series
- Mistral™ Series
- Phi™ Series
- Gemma™ Series

Ideal for:

- AI Assistants
- Enterprise Chatbots
- Local LLM Deployment
- RAG Applications

Speech & Audio AI

- Whisper Large v3

Ideal for:

- Meeting Transcription
- Voice Assistants
- Speech Analytics

Vision Language Models (VLMs)

- Qwen2.5-VL
- Qwen-VL Series
- InternVL Series

Ideal for:

- Multimodal AI
- Document Understanding
- Visual QA
- Intelligent Search

OCR & Document AI

- DeepSeek OCR
- PaddleOCR-VL

Ideal for:

- Enterprise Document Processing
- Invoice Analysis
- Knowledge Base Creation



UGen Intel® Arc™ Pro B70 32GB

UGEN-ARC-PRO-B70-32G

Ubiquitous AI. Powering Ingenuity.

Specifications

Product Name		UGen Intel® Arc™ Pro B70 32GB
Model name		UGEN-ARC-PRO-B70-32G
GPU		X ^e 2 Architecture
X ^e -Core		32
Render Slice		8
Ray Tracing Units		32
Intel® X ^e Matrix Extensions (Intel® XMX) Engines		256
X ^e Vector Engines		256
Max. Boost Clock		2800 MHz
AI Performance	TFLPOS(FP32)	22.94
	pTOPS	367
TBP		290 W
PCI Express Configurations		PCI Express 5.0 x16
Power Cable		12V-2x6
Memory		32 GB GDDR6
Graphics Memory Interface		256 bit
Graphics Memory Bandwidth		608 GB/s
ECC Memory Supported		Yes
# of Displays Supported		4
Graphics Output		3x DisplayPort 2.1, 1x HDMI
Max Resolution		7680 x 4320@120Hz
H.264 Hardware Encode/Decode		Yes
H.265 (HEVC) HW Encode/Decode		Yes
AV1 Encode/Decode		Yes
Ray Tracing		Yes
oneAPI Support		Yes
OpenVINO™ Support		Yes
DirectX* Support		DirectX 12 Ultimate
Vulkan* Support		1.3
OpenGL* Support		Up to 4.6
OpenCL* Support		3.0
Media Profiles		HDR10, HDR10+ GAMING, Dolby Vision
Multi-Format Codec Engines		2
VESA Adaptive Sync		Yes
Intel® Extension for PyTorch* (IPEX) Support		Yes
Intel® X ^e Super Sampling (XeSS) Support		Yes
Card Size (mm)		267 mm x 111 mm x 40 mm
Weight (Kg)		1.094
Form Factor		2-slot
Certifications		■ FCC ■ CE ■ RCM ■ BSMI ■ VCCI ■ UKCA ■ KCC ■ UkrCEP



UGen Intel® Arc™ Pro B70 32GB
UGEN-ARC-PRO-B70-32G

Ubiquitous AI. Powering Ingenuity.

Photos

