

## What if Your PC Does Not Have an NPU?

364 2:36

If you don't have an NPU, your AI tasks will fall back to the GPU first, then the CPU. That's why NPUs are being added – they're more power-efficient than GPUs for continuous, lightweight AI tasks (like live video filters, speech recognition, or background AI assistants).

### GPU vs. NPU Performance on Laptops

#### 1. Raw AI Power

**GPU** - Much higher peak throughput (teraflops). Can handle big models (LLMs, Stable Diffusion, video upscaling). Example: Nvidia RTX 4070 Laptop GPU → ~100-150 TOPS (AI ops).

**NPU** Lower raw power (~10-45 TOPS today on Intel Core Ultra or AMD Ryzen AI). Designed for specific AI inference tasks (image filters, noise suppression).

#### 2. Efficiency (Performance per Watt)

**GPU**- Power-hungry – running AI inference can pull 30-80W+ on a laptop. Drains battery quickly if used for continuous background tasks.

**NPU** - Very efficient – usually 1-5W for AI tasks. Ideal for always-on features like live captions, background blur, or voice isolation without killing battery life.

#### 3. Task Suitability

**GPU Best For:** - Training or running large AI models locally (LLMs, Stable Diffusion). Heavy workloads (video rendering, 3D, scientific compute).

**NPU Best For:** - Lightweight, constant AI tasks (real-time transcription, camera effects, personal assistant inference). Offloading from GPU/CPU so battery lasts longer.

#### 4. Integration in Laptops

**GPU** - Discrete GPUs (Nvidia RTX, AMD Radeon) are still required for high-end creative/AI workloads. Integrated GPUs (Intel Xe, AMD Radeon iGPU) can also run AI, but less powerful.

**NPU** - Integrated into CPUs (Intel Core Ultra, AMD Ryzen AI, Apple M-series). Works alongside GPU/CPU, automatically chosen by Windows or macOS for certain tasks.

## 5. Real-World Examples

- Adobe Photoshop (Generative Fill): GPU preferred (big models).
- Zoom Background Blur: NPU preferred (low-power continuous AI).
- Stable Diffusion image generation: GPU required (too heavy for NPU).
- Windows Studio Effects (eye contact, noise canceling): NPU handles it efficiently.

### Recap

- GPU = muscle → raw AI power, big models, creative workloads.
- NPU = brainstem → small, efficient, always-on AI tasks that run in the background.
- On modern laptops, they complement each other: the OS decides which processor to use depending on the workload.