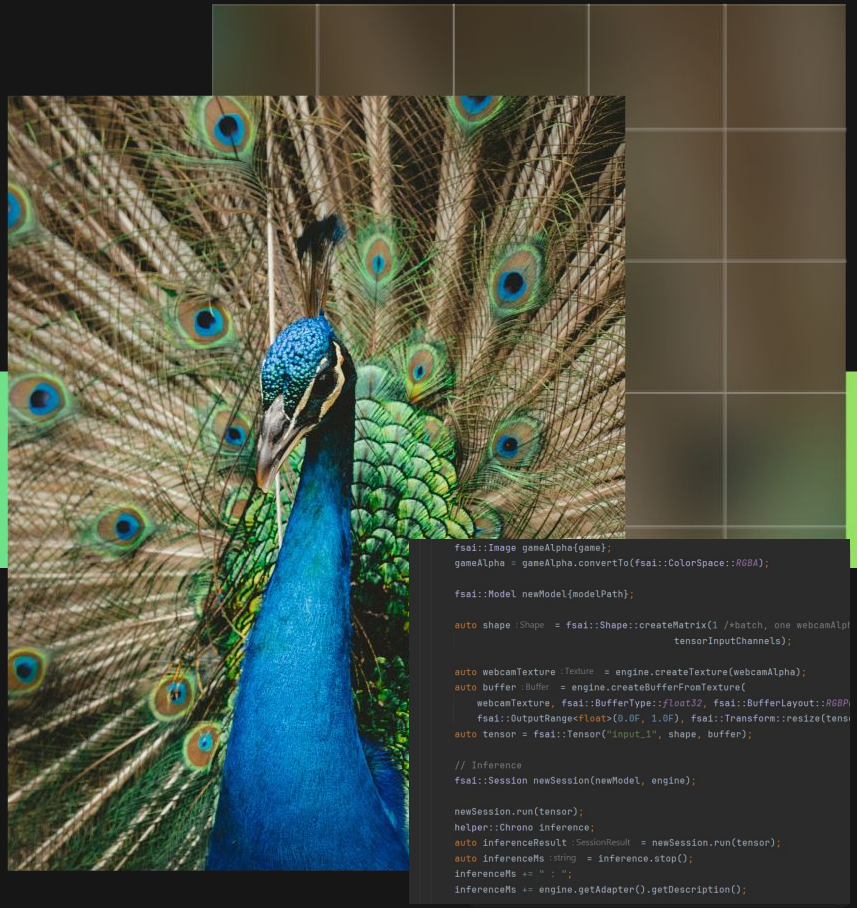
fluendo

Full GPU driven AI workloads with GStreamer and Raven



Andoni Morales
Nacho García
Aleix Figueres
Izan Leal
Sergio Sánchez

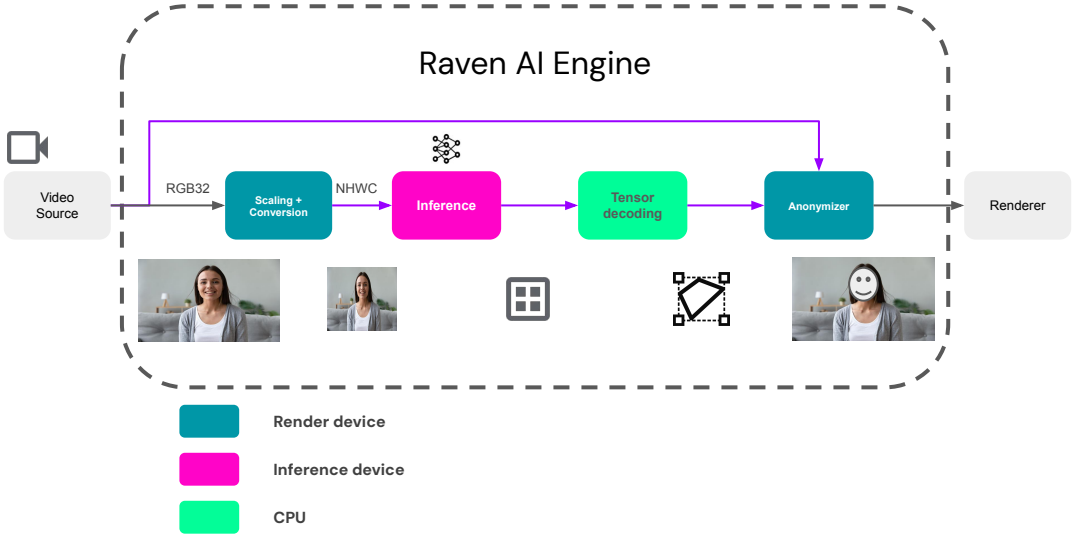
GStreamer Conference 2025

Raven AI Engine

Raven is our in-house AI engine designed for real-time multimedia workflows, combining high-throughput AI inference with graphics-native processing.

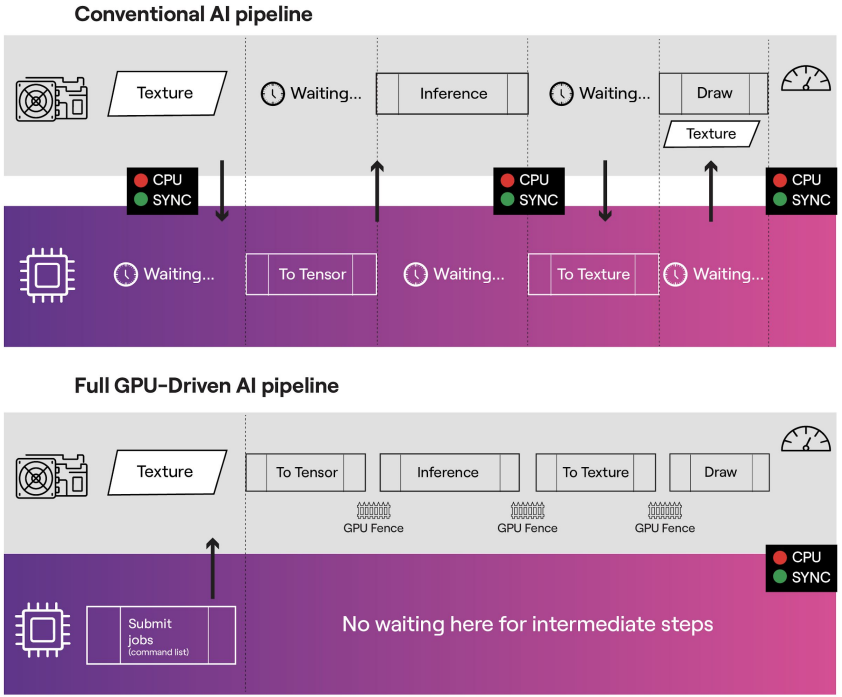
It gives full control over the entire GPU pipeline, from memory allocation to execution, allowing for deep customization across hardware and environments.

- 100% GPU driven processing
- Multivendor compatibility
- Multiplatform support
- Multipurpose design
- Low code integration



Video

Full GPU-Driven AI pipeline



- Task parallelization
- Post-processing algorithms on the GPU (eg: Non-Max Supression)
- GPU rendering engine

Anonymization task runs at 500 fps for 4K (3840x2160) input a in an Nvidia RTX 4060 Desktop.