



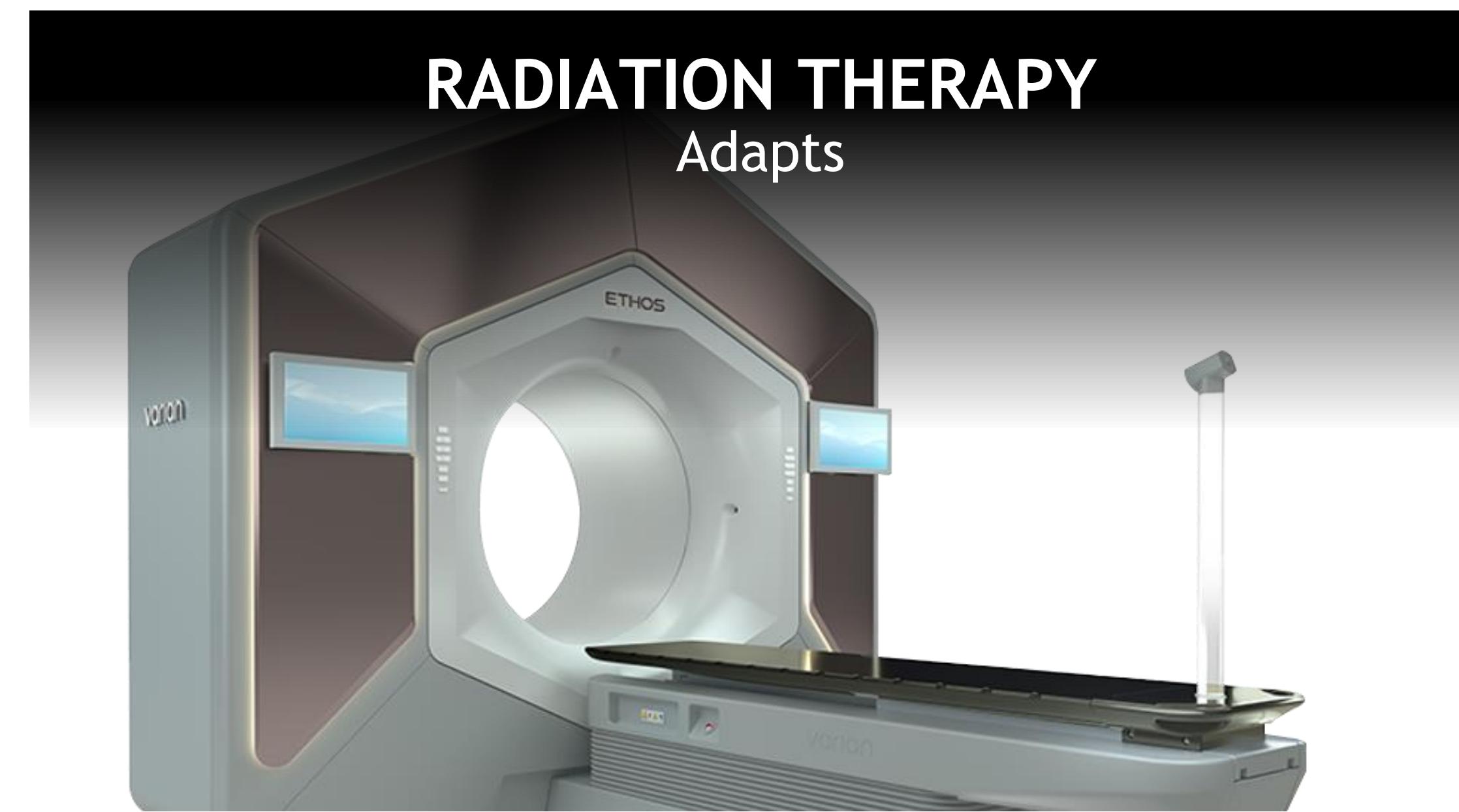
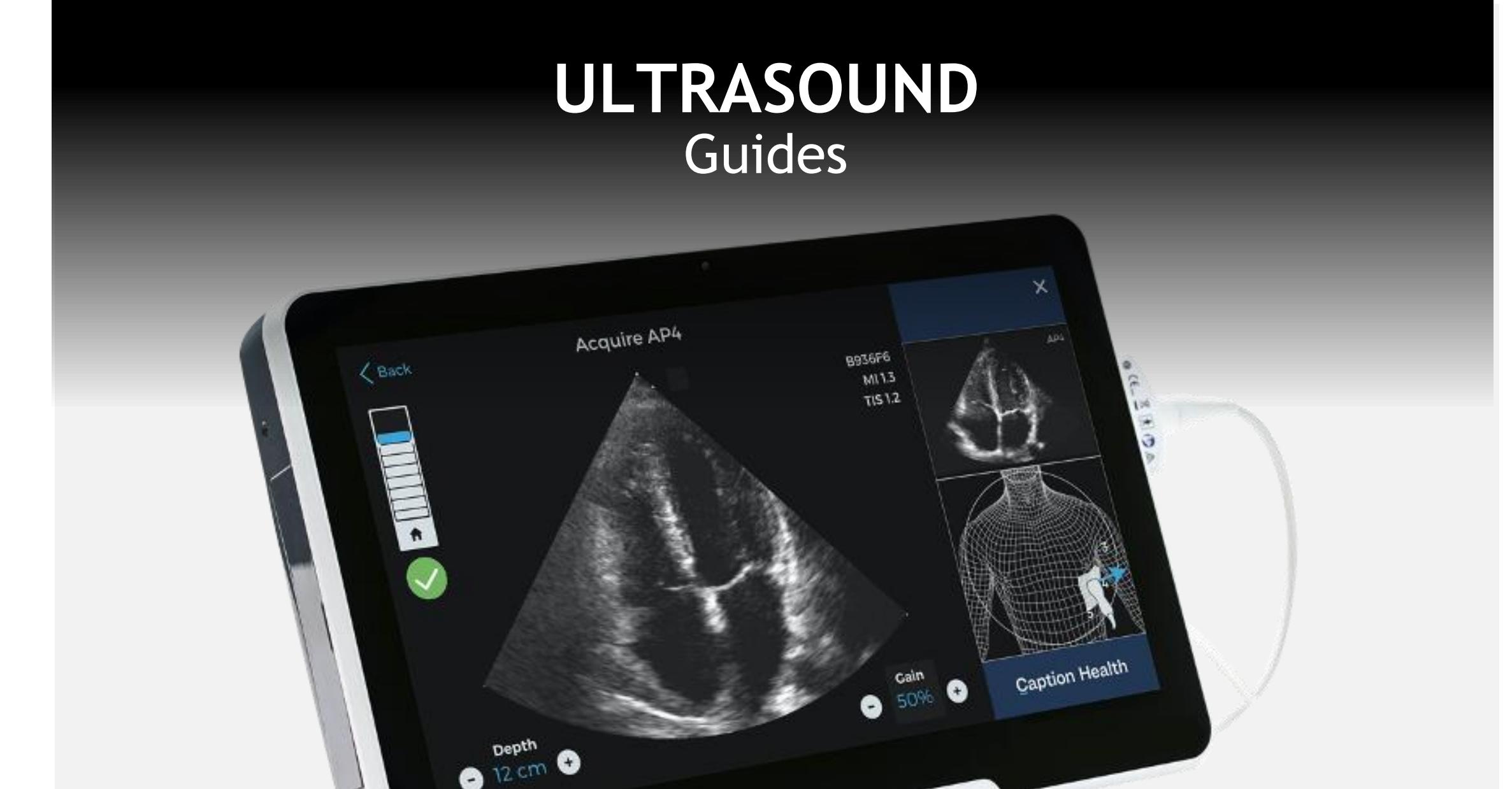
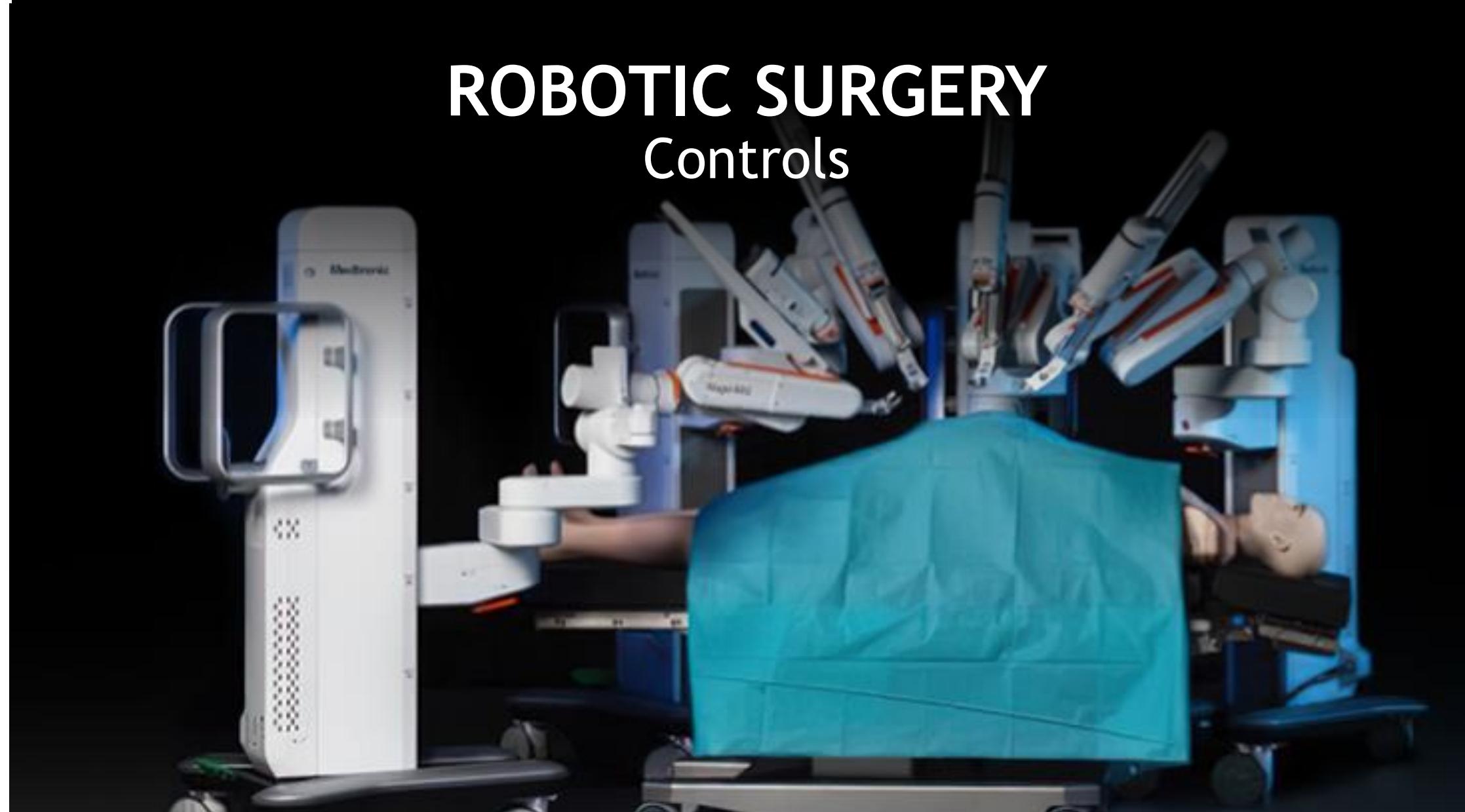
# Clara Holoscan

AI Computing Platform for Medical Devices



# Every Medical Device Will Become Robotic

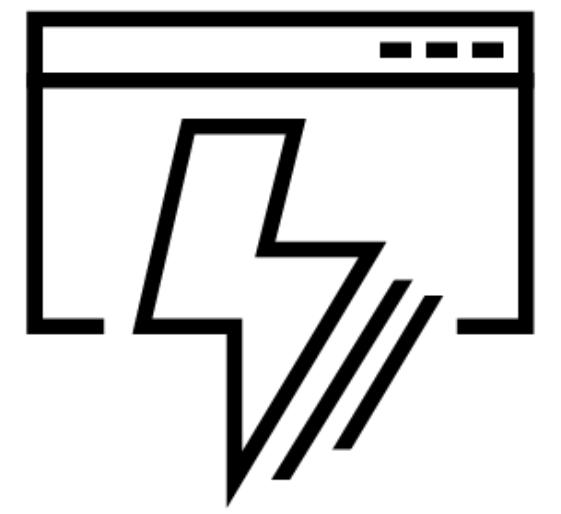
Performing AI Actions or Skills in Real-time



# Medical Device Developers Need an AI Computing Platform

## Challenges in Medical Device Development

### Developer Productivity



Embedded & FPGA Dev

Sensor IO

Performance

### Security



Cybersecurity

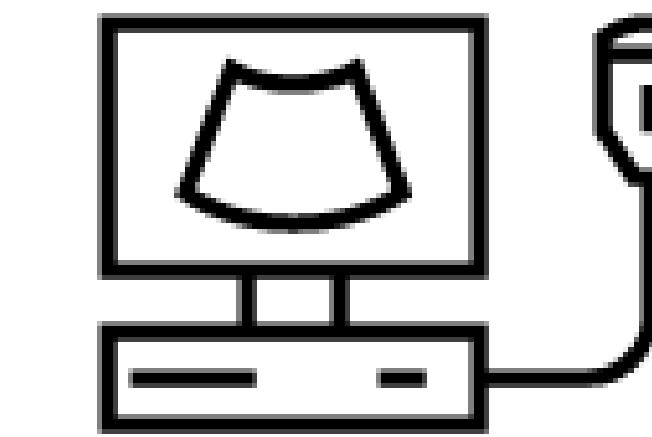
Remote provisioning and management

### Medical Grade



Regulatory Requirements  
(IEC 60601, 62304).

### Production Deployment



Production Hardware

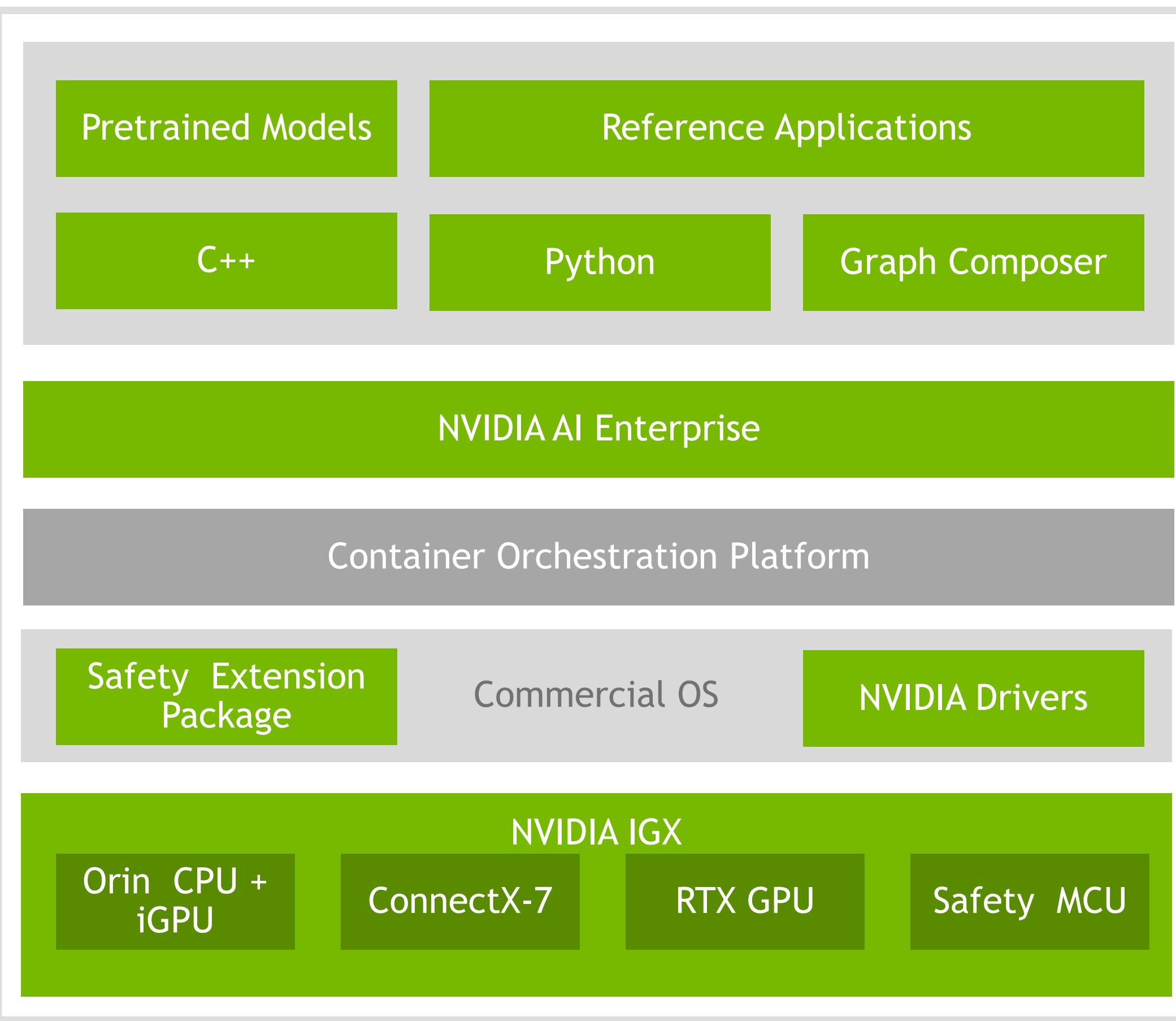
Extended Support

# NVIDIA Clara Holoscan Platform

## AI Computing Platform for Medical Devices

### Clara Holoscan SDK

#### Develop



### Developer Kits

#### Validate



NVIDIA IGX Orin DevKit (EA)  
Orin, RTX A6000, ConnectX-7  
Available early 2023



Clara AGX DevKit (GA)  
AGX Xavier, RTX 6000, ConnectX-6  
Available Now

### NVIDIA IGX

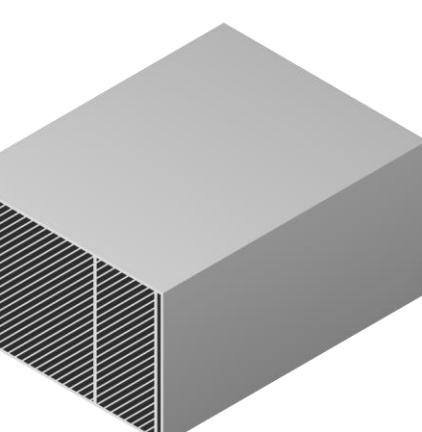
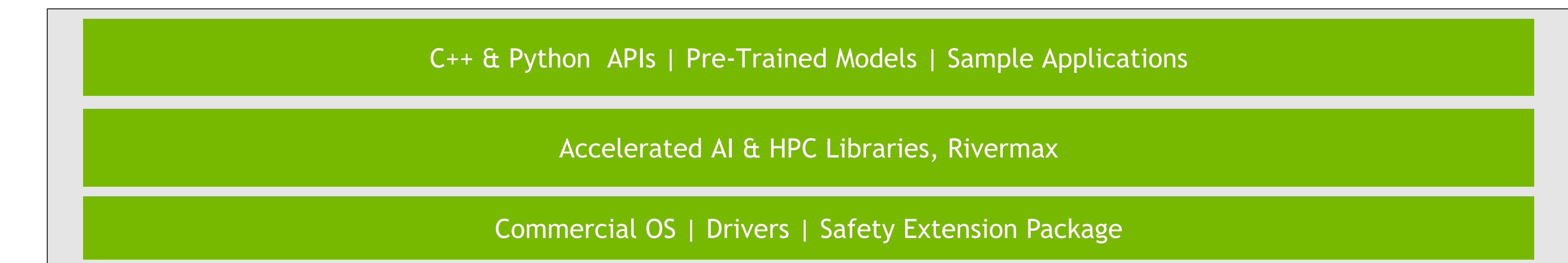
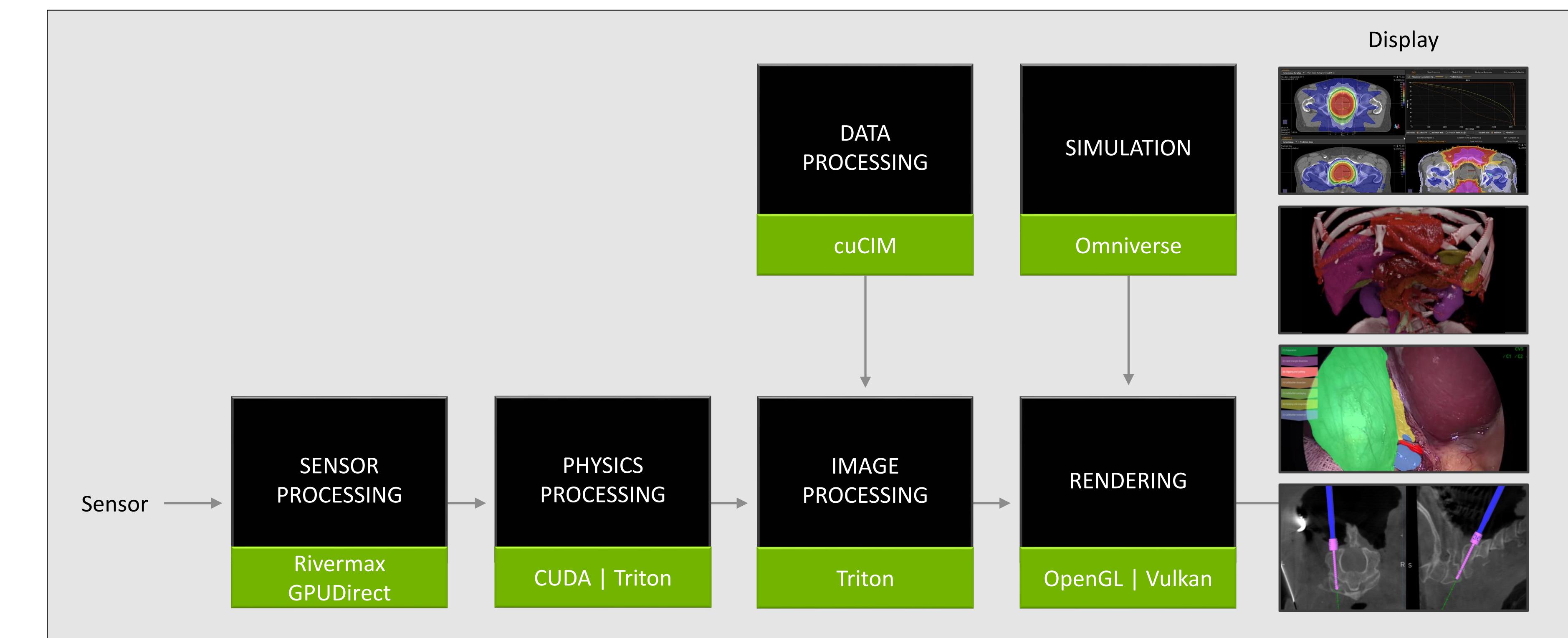
#### IEC 60601 Ready



# NVIDIA Clara Holoscan Platform

## AI Computing Platform for Medical Devices

- Optimized for Streaming AI
  - Accelerated AI and HPC Libraries
  - Rivermax for GPUDirect RDMA Data Streaming
  - Pre-trained Models, Sample Applications (C++, Python)
- Safety, Security and Manageability Built In
  - Safety Extension Package
  - Functional Safety Island
  - sMCU
- Built for Medical Certification (IEC 60601, 62304)
- Long life Hardware & Long-term Software Support
- Rich sensor partner ecosystem



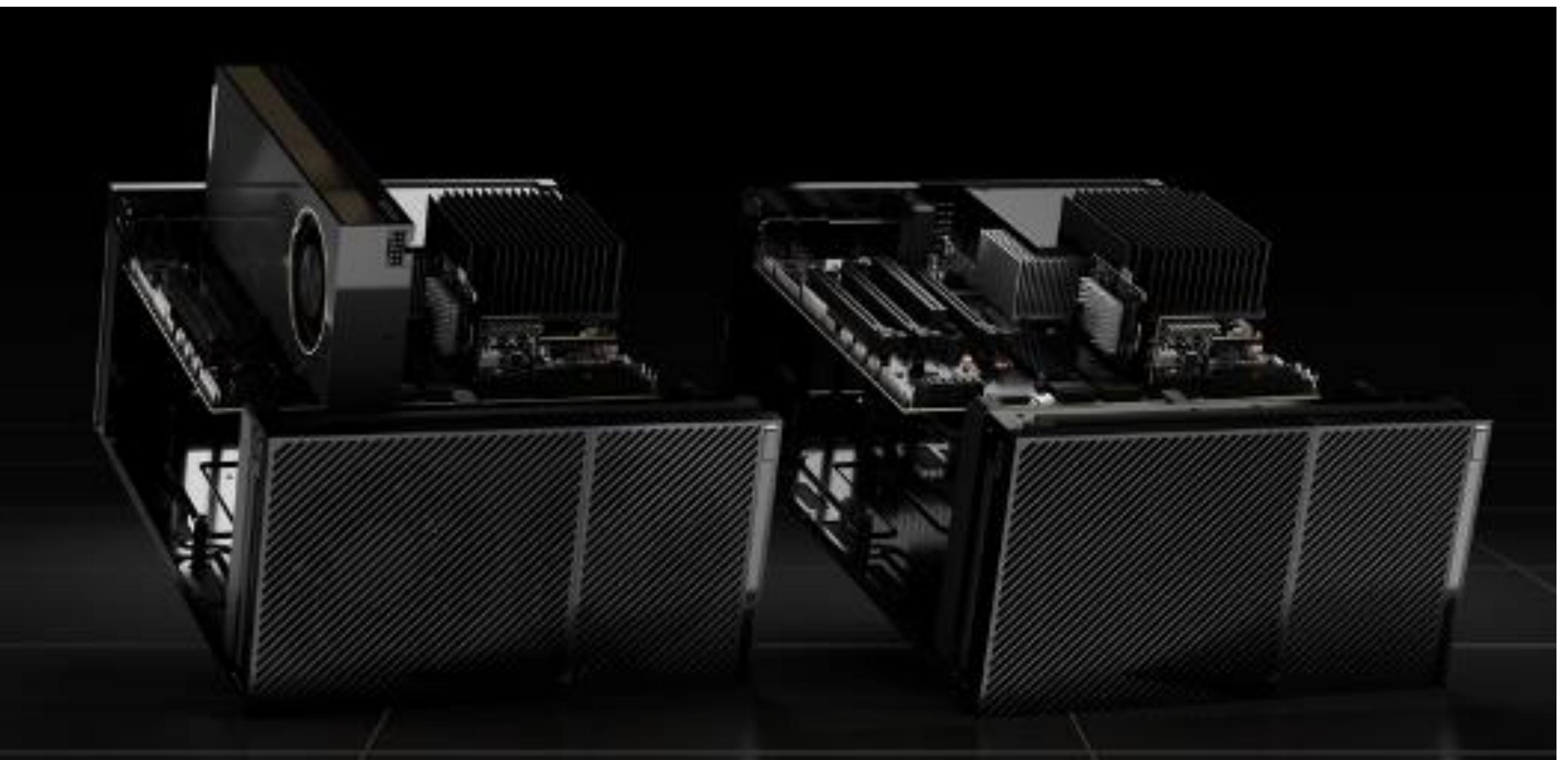
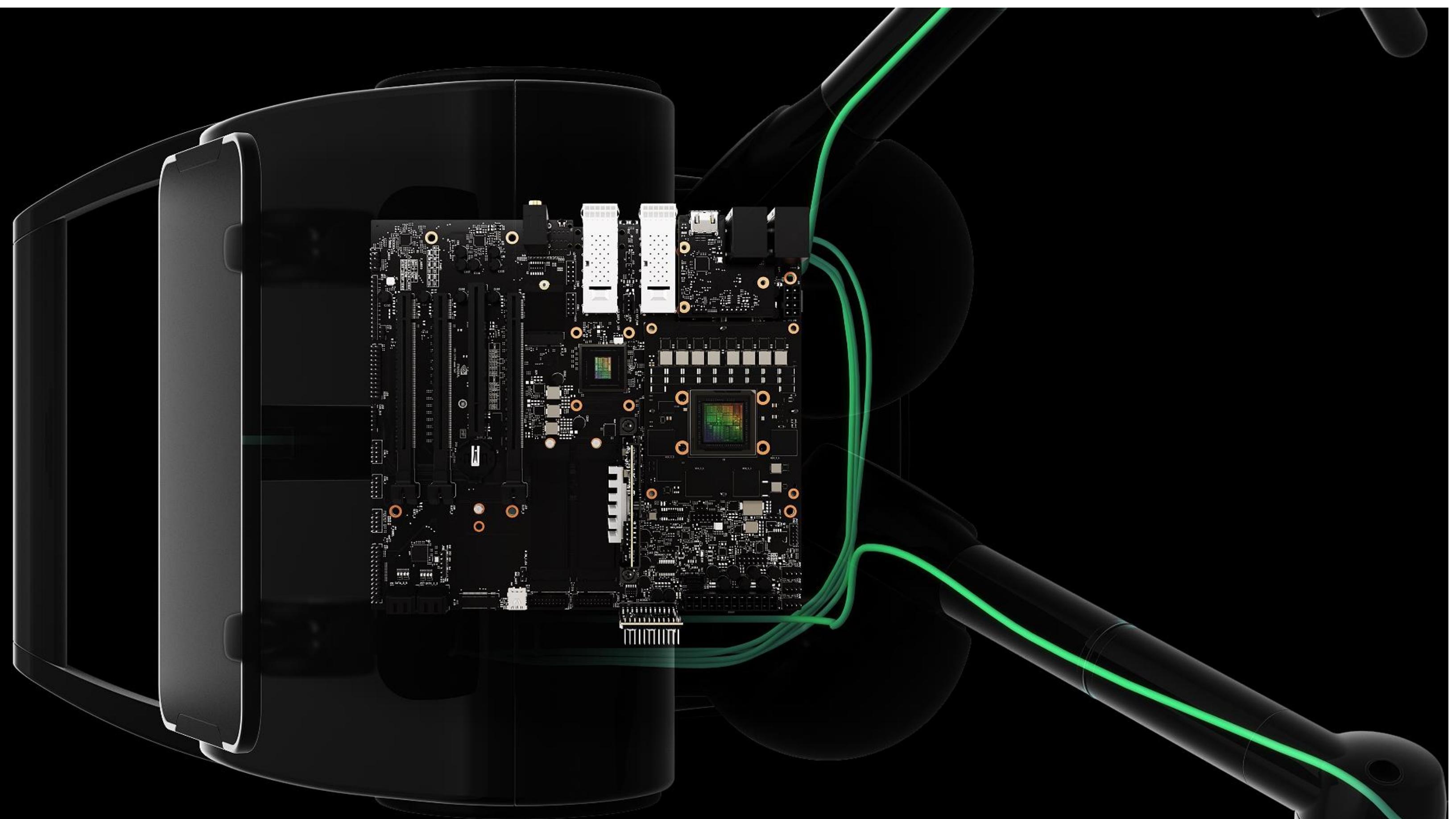
**IGX**  
Orin, ConnectX-7, A6000  
Functional Safety Island, Safety MCU

# Clara Holoscan on NVIDIA IGX

Deploying real-time AI apps to clinical production

Available early 2023

- Orin 12 Core ARM System on Chip, 250 TOPS FP8, 64GB
- RTX A6000 Tensor Cores, RT Cores, CUDA Cores, 600 TOPS FP8
- ConnectX-7 High Speed Edge Networking 2x 100GbE
- Functional Safety Island & Safety MCU
- Rich I/O HDMI | PCIe | Ethernet | WiFi
- Scalable AI Architecture 250-600 TOPS
- Sensor Front End Processing 4K SDI | CoaXpress | Ultrasound
- DevKits for prototyping and validation
- Production hardware through medical embedded OEMs
- Built for Medical IEC 60601



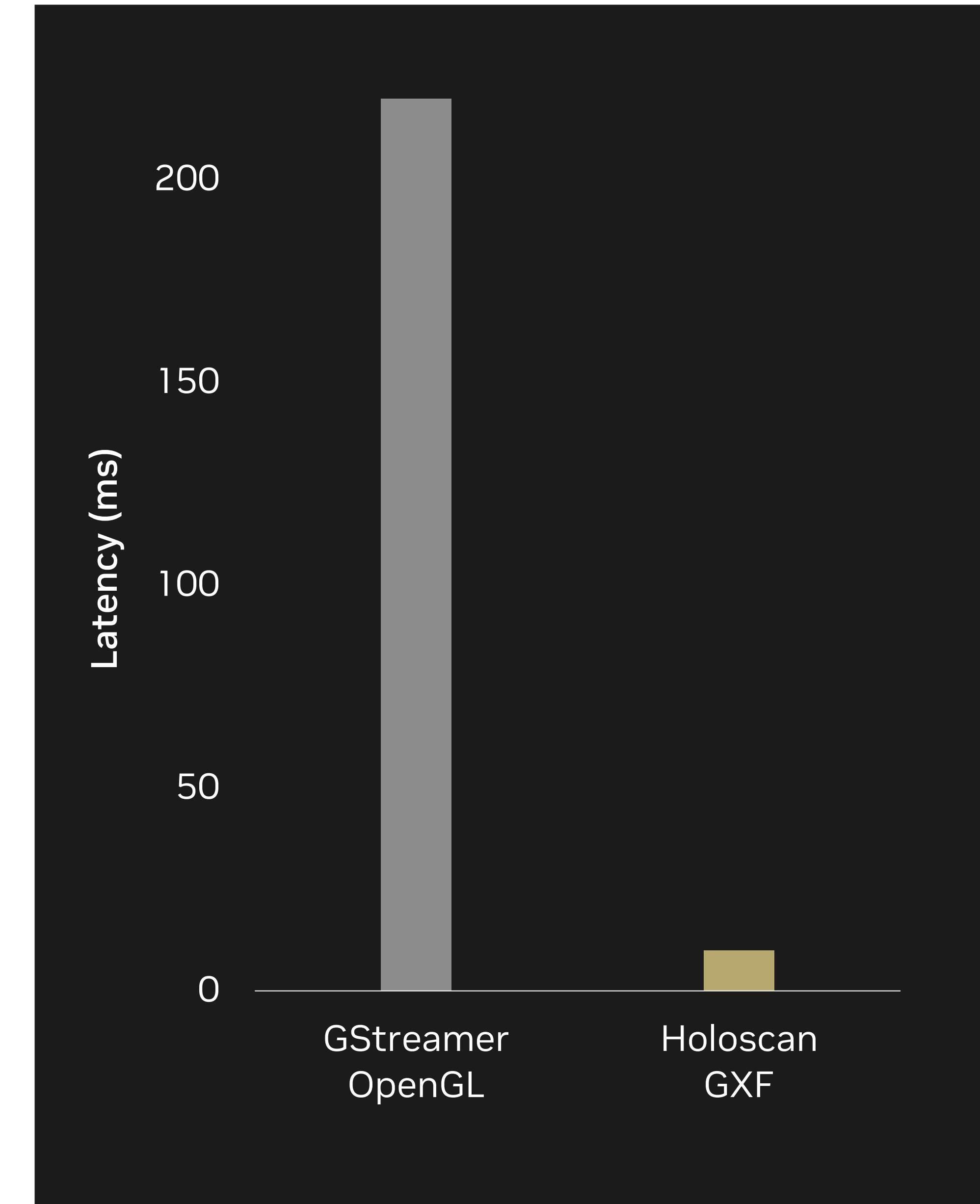
# NVIDIA Clara Holoscan

## Developer Productivity & Performance

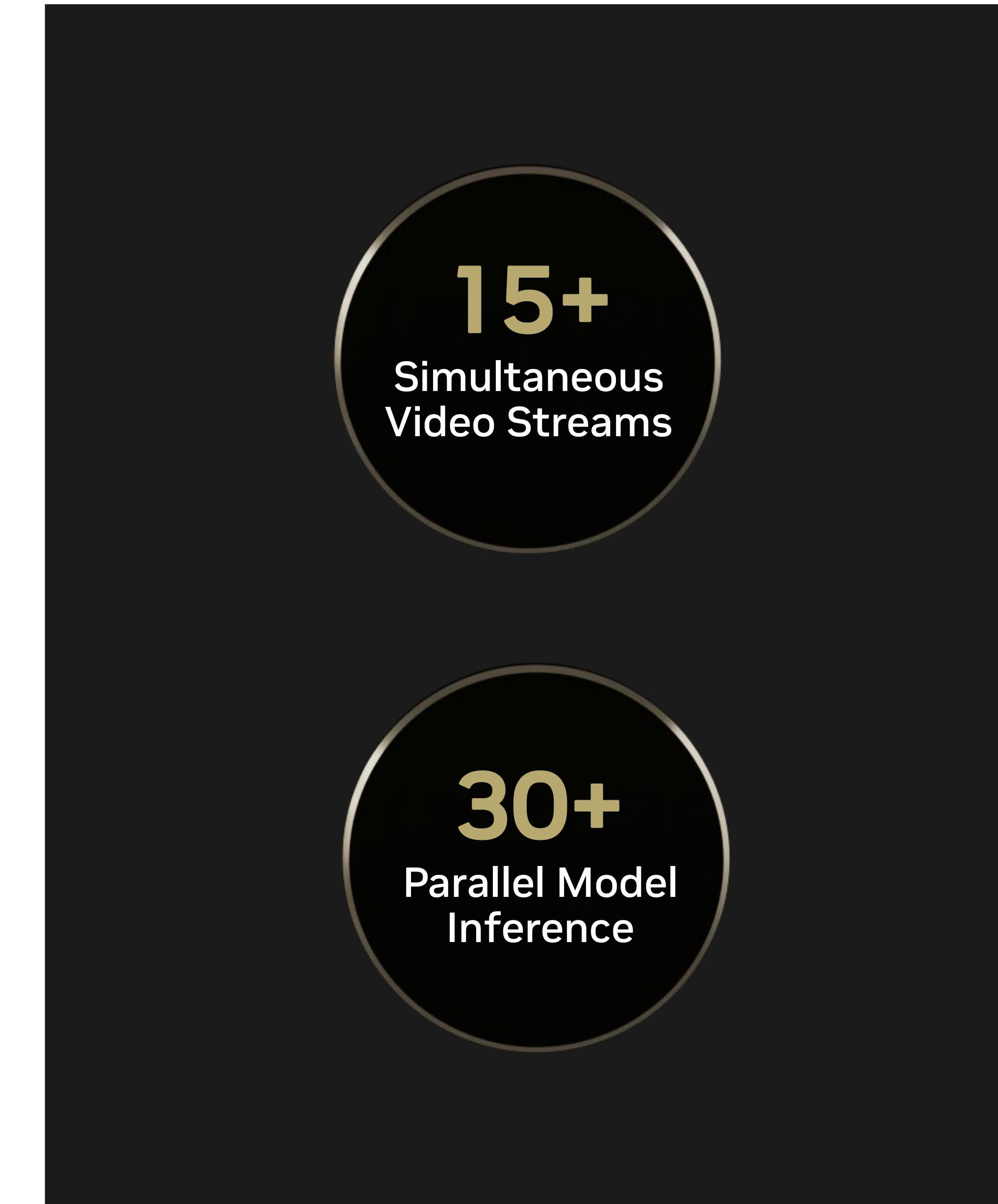
MULTI-MODALITY SENSORS SUPPORT  
Video, Ultrasound, Laser



SUPER LOW LATENCY  
4K 240Hz Video Processing Pipeline



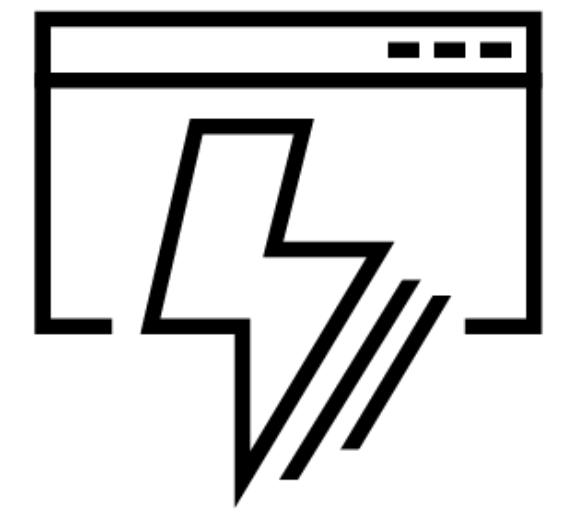
ULTIMATE EDGE AI PERFORMANCE  
Stream AI Processing <50ms



# Holoscan provides a single AI computing platform

Accelerating time to market for next gen medical devices

## Developer Productivity



Optimized for High-performance

Sensor Partner Ecosystem

C++, Python (coming soon)

## Secure



Secure by design

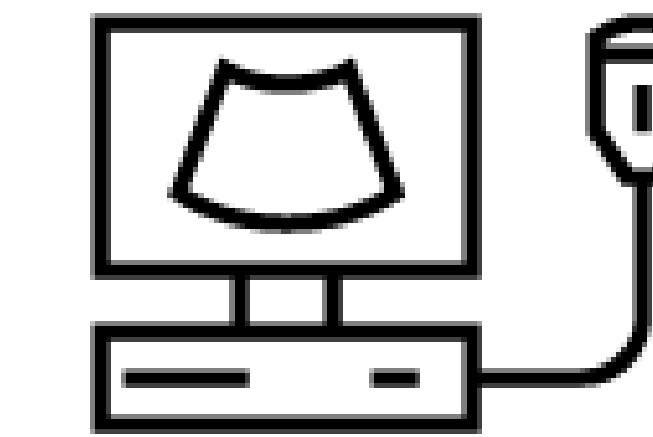
Remote provisioning and  
management

## Medical Grade



Ready for certification  
(IEC 60601, 62304).

## Production Ready

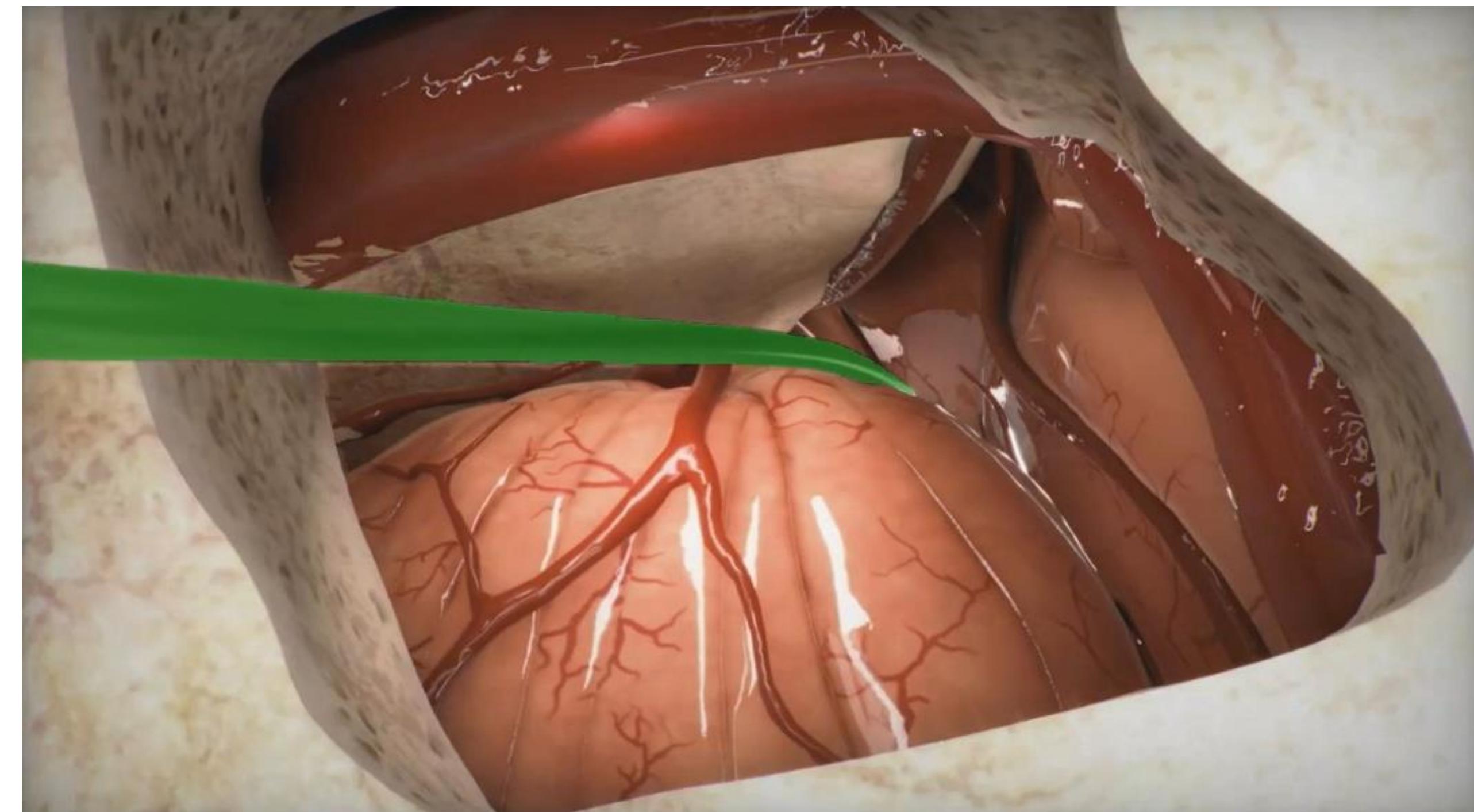
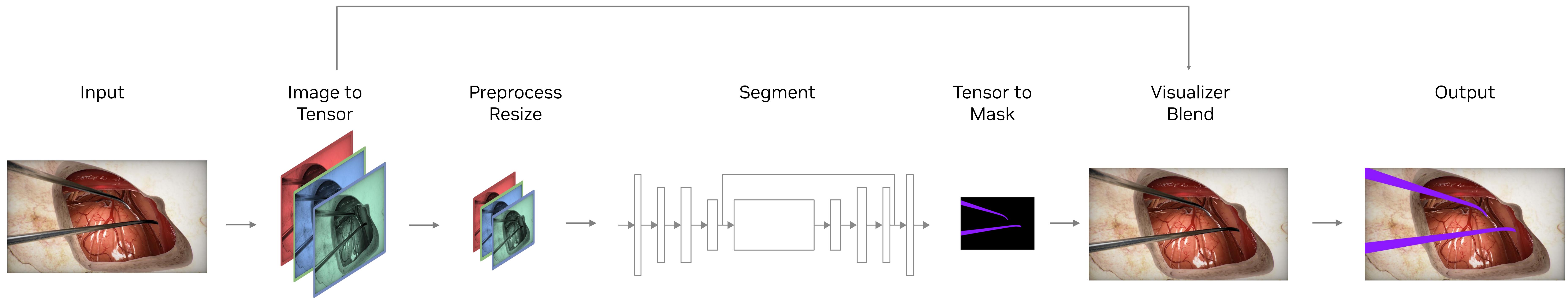


Customizable white-label  
platform

Long term support

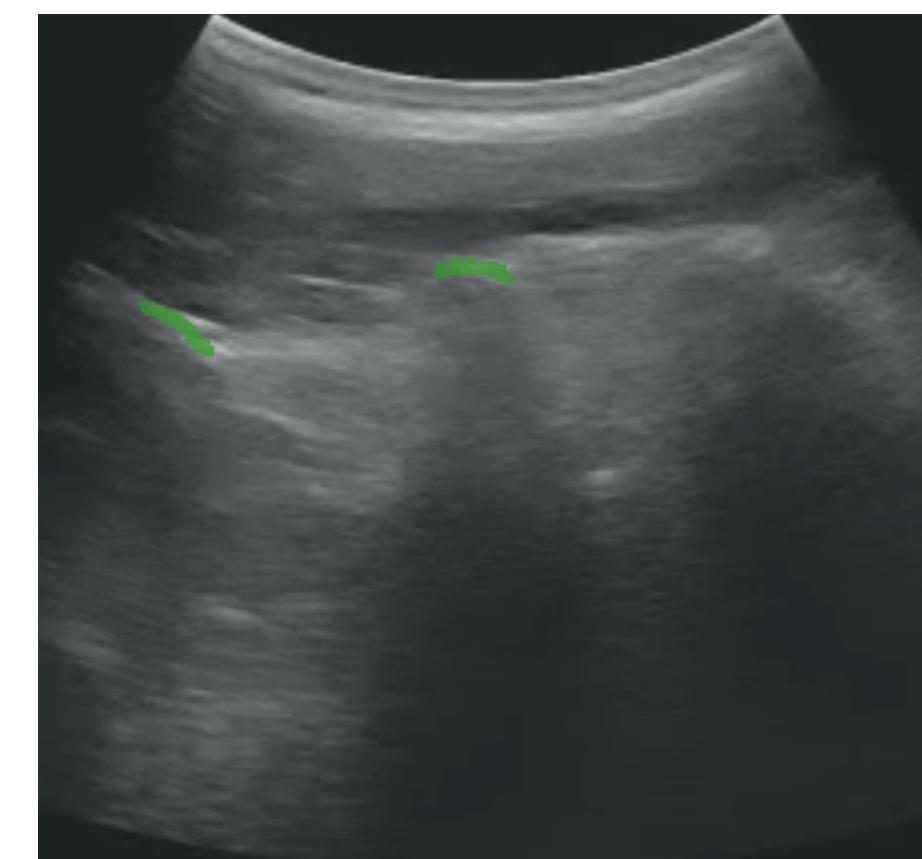
# Clara Holoscan SDK

## Endoscopy AI Pipeline



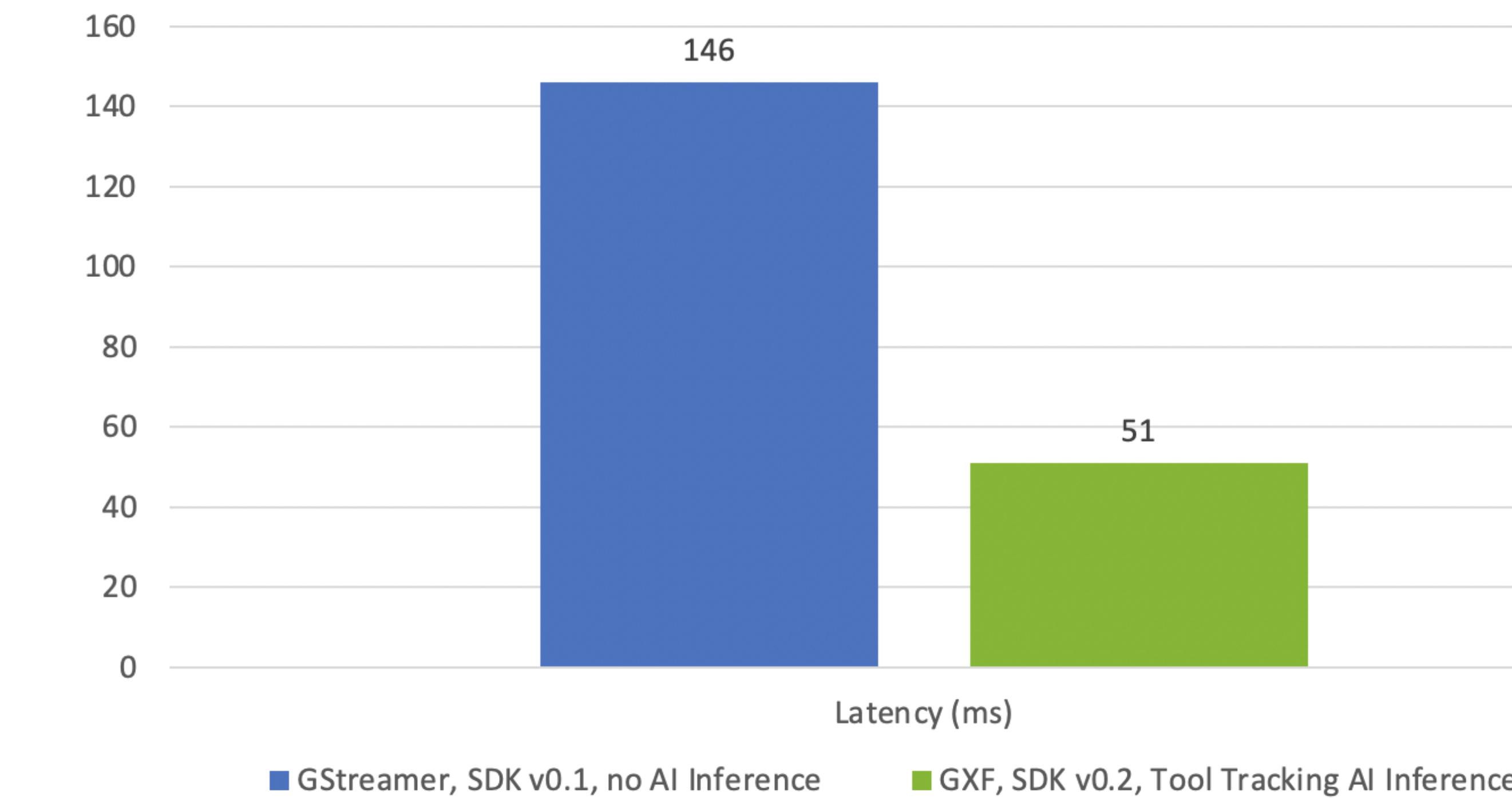
# Clara Holoscan SDK

## Endoscopy & Ultrasound Sample Apps



Endoscopy Tool Tracking  
Ultrasound Segmentation  
BYOM: Bring your own Model

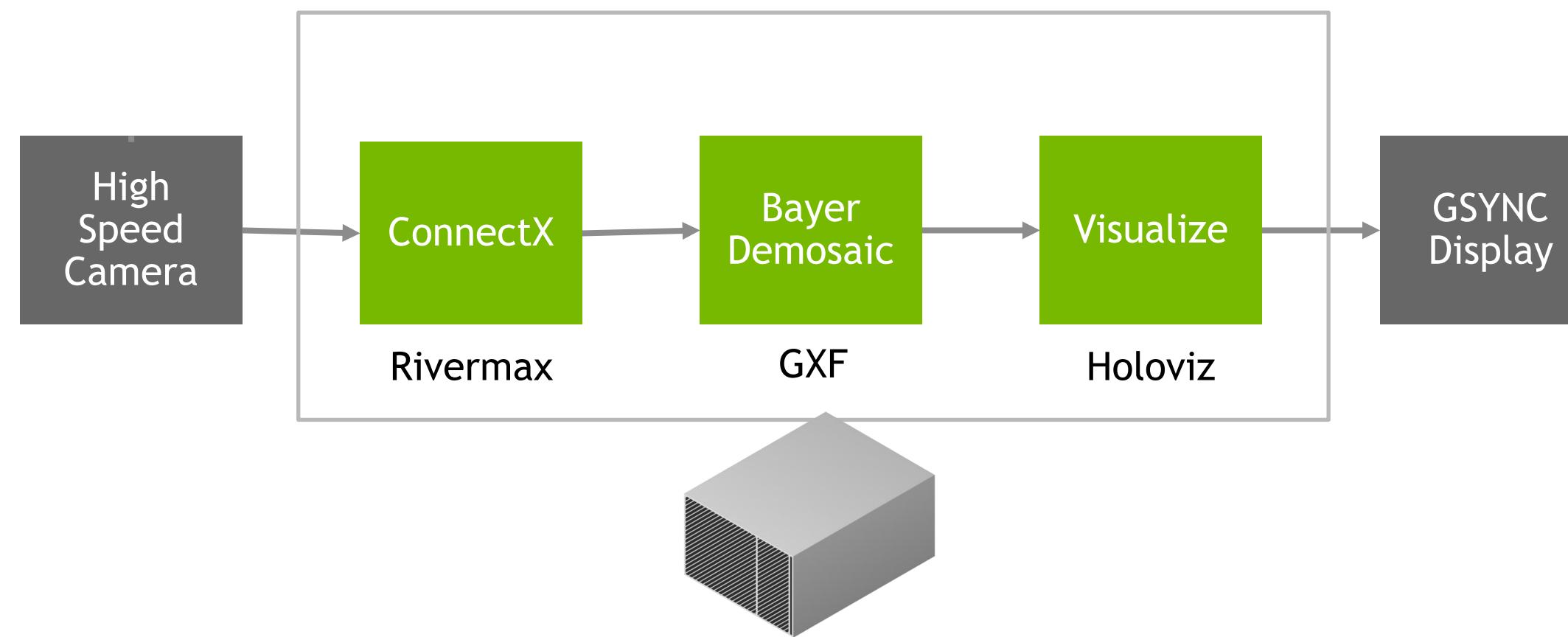
## Optimized for Streaming AI



3x Reduction in Processing Overhead  
Compared to GStreamer

# Introducing Clara Holoscan 0.3

## High Speed Endoscopy (4K 240Hz)



10ms End-to-End Latency

## C++ APIs

```
// Flow definition
add_flow(replayer, visualizer_format_converter);
add_flow(visualizer_format_converter, visualizer,
         {"tensor", "source_video"});
add_flow(replayer, format_converter);
add_flow(format_converter, lstm_inferer);
add_flow(lstm_inferer, visualizer,
         {"tensor", "tensor"});
add_flow(replayer, recorder);
```

Improved Developer Experience

## x86 Support



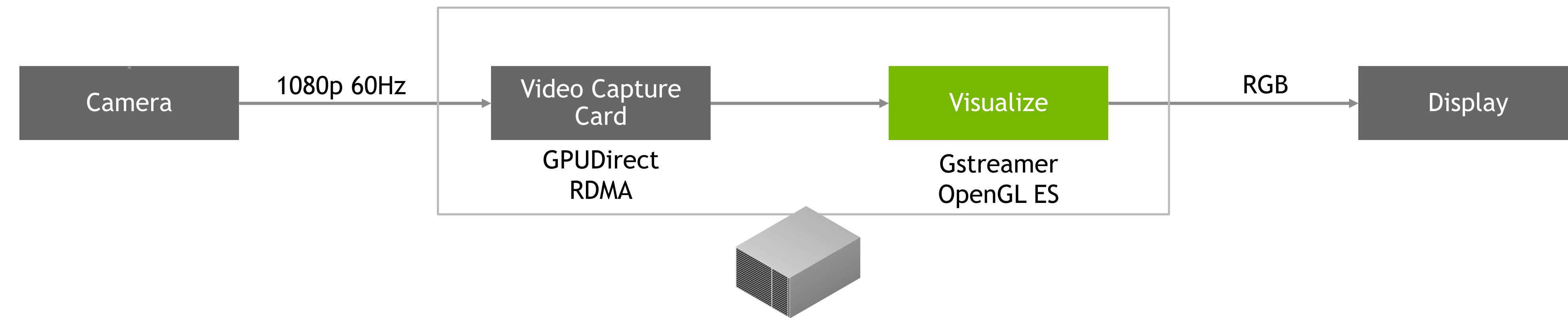
Prototype on Laptops and Workstations

# High Speed Endoscopy

## 4K 240FPS

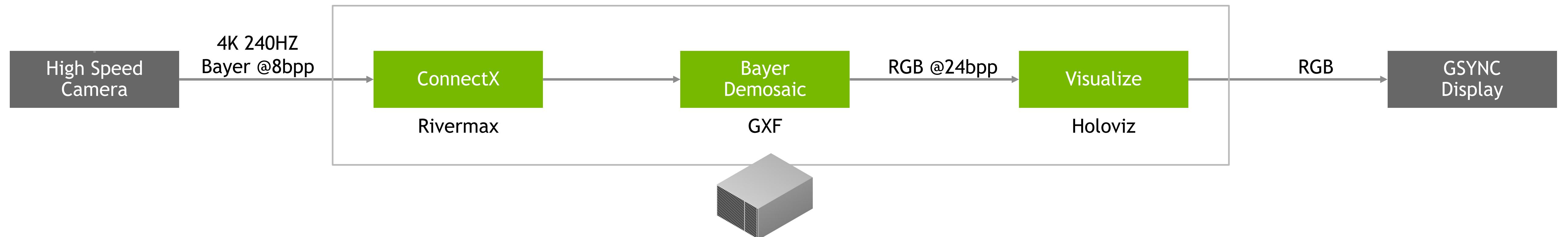
**Traditional Endoscopy pipeline**  
(GStreamer, OpenGL)

**220ms** End-to-End Latency  
for 1080p 60Hz



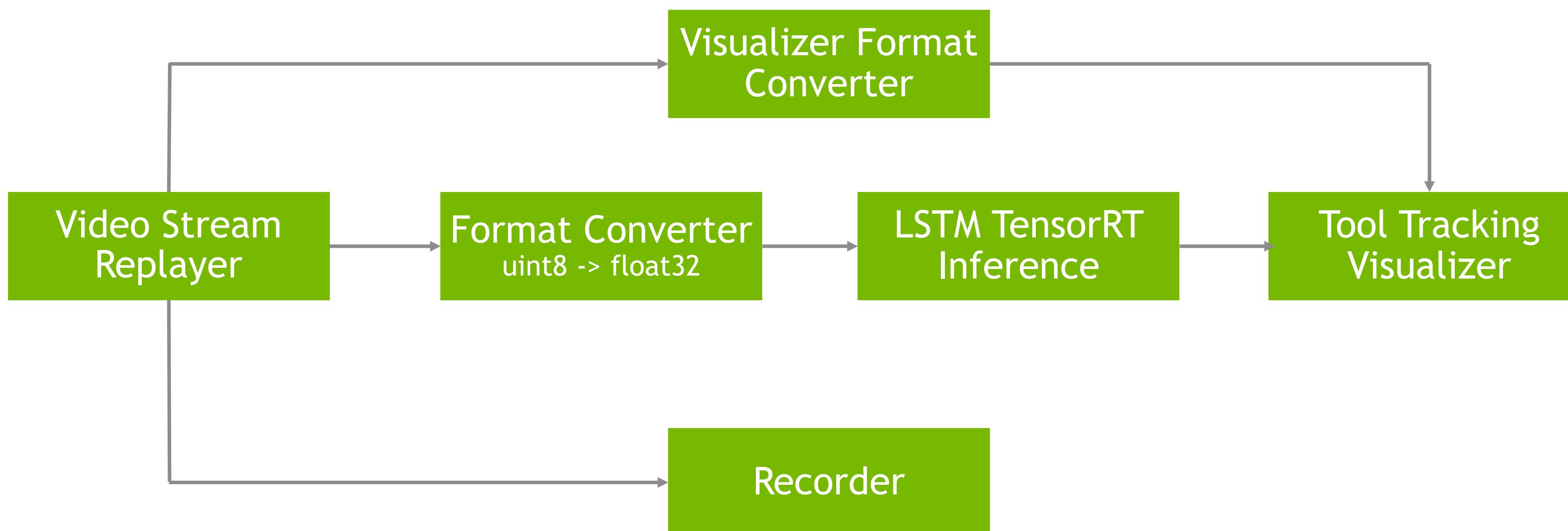
**High Speed Endoscopy Pipeline**  
(Rivermax, GXF, GSync)

**10ms** End-to-End Latency  
for 4K 240Hz



# New C++ APIs

## Improved Developer Experience



```
class App : public holoscan::Application {
public:
    void compose() override {
        using namespace holoscan;

        auto replayer = make_operator<ops::VideoStreamReplayerOp>("replayer", from_config("replayer"));

        auto recorder = make_operator<ops::VideoStreamRecorderOp>("recorder", from_config("recorder"));

        auto format_converter = make_operator<ops::FormatConverterOp>(
            "format_converter",
            from_config("format_converter_replayer"),
            Arg("pool") = make_resource<BlockMemoryPool>("pool", 1, 4919041, 2));

        auto lstm_inferer = make_operator<ops::LSTMTensorRTInferenceOp>(
            "lstm_inferer",
            from_config("lstm_inference"),
            Arg("pool") = make_resource<UnboundedAllocator>("pool"),
            Arg("cuda_stream_pool") = make_resource<CudaStreamPool>("cuda_stream", 0, 0, 0, 1, 5));

        auto visualizer_format_converter = make_operator<ops::FormatConverterOp>(
            "visualizer_format_converter",
            from_config("visualizer_format_converter_replayer"),
            Arg("pool") = make_resource<BlockMemoryPool>("pool", 1, 6558720, 2));

        auto visualizer = make_operator<ops::ToolTrackingVizOp>(
            "visualizer",
            from_config("visualizer"),
            Arg("pool") = make_resource<UnboundedAllocator>("pool"));

        // Flow definition
        add_flow(replayer, visualizer_format_converter);
        add_flow(visualizer_format_converter, visualizer, {"tensor", "source_video"});

        add_flow(replayer, format_converter);
        add_flow(format_converter, lstm_inferer);
        add_flow(lstm_inferer, visualizer, {"tensor", "tensor"});

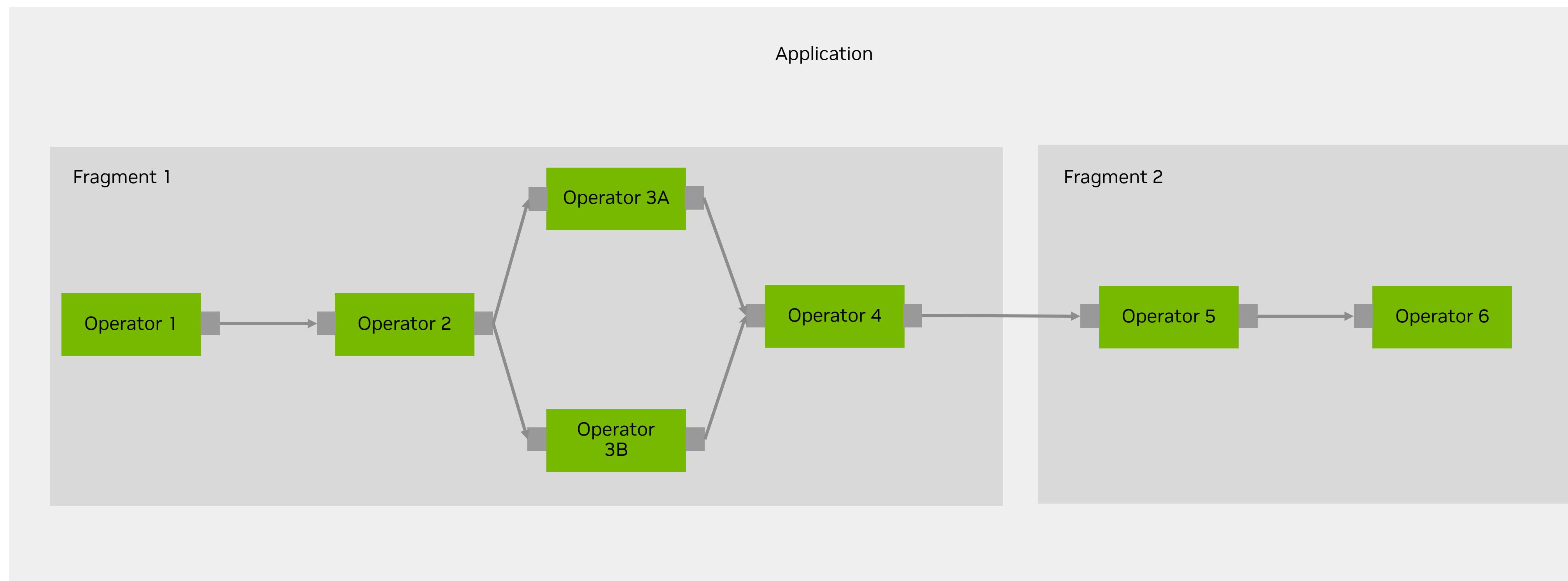
        add_flow(replayer, recorder);
    }
};

int main() {
    App app;
    app.config("apps/endoscopy_tool_tracking/app_config.yaml");
    app.run();

    return 0;
}
```

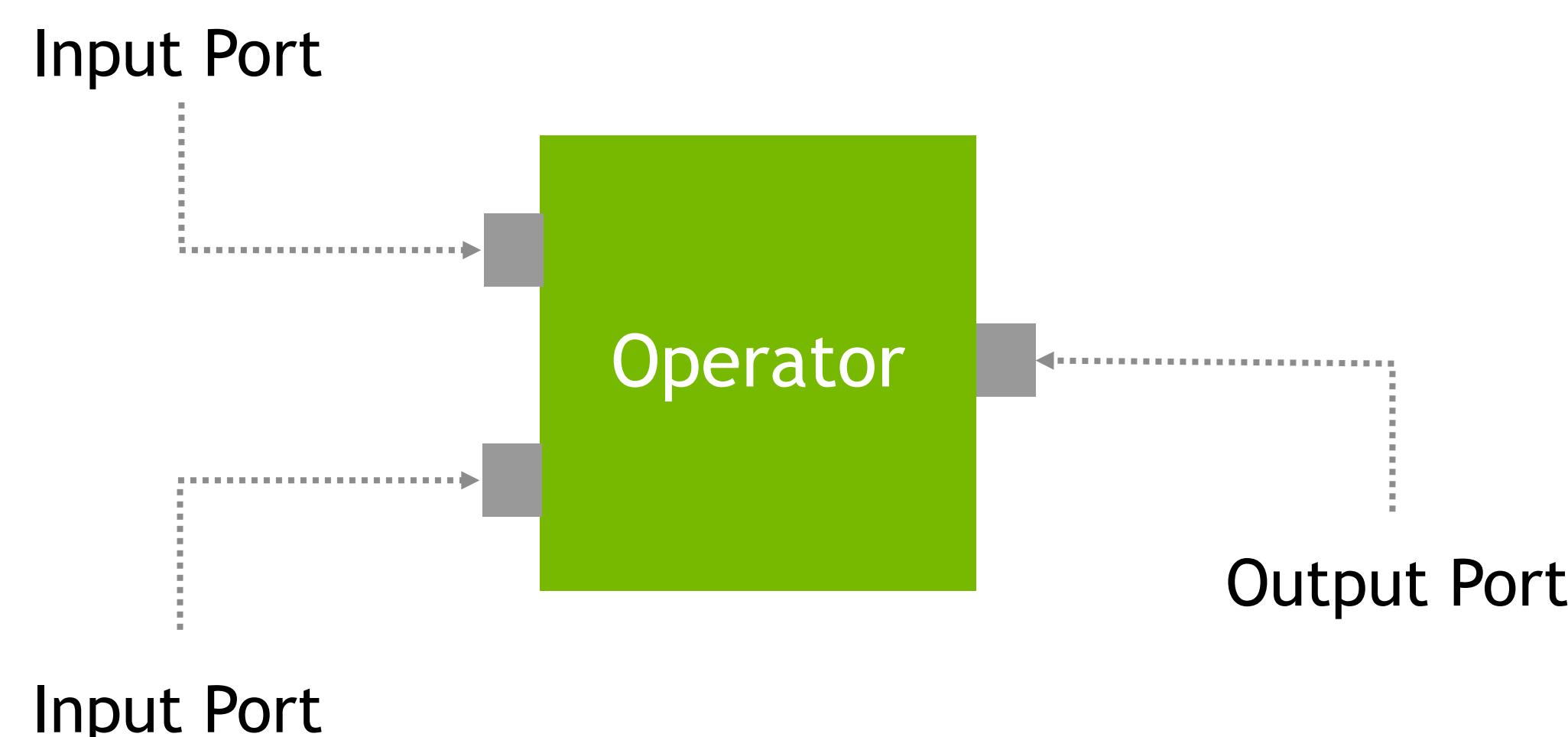
# C++ APIs

## Architecture



**Application** acquires and processes streaming data. It's a collection of fragments where each fragment can be allocated to execute on a physical node of a Holoscan cluster.

**Fragment** is a Directed Acyclic Graph (DAG) of operators. It can be assigned to a physical node of a Holoscan cluster during execution. The run-time execution manages communication across fragments. In a Fragment, Operators (Graph Nodes) are connected to each other by flows (Graph Edges).



**Operator** is the most basic unit of work. It receives streaming data at an input port, processes it, and publishes it to one of its output ports.

**Port** is an interaction point between two operators. Operators ingest data at Input ports and publish data at Output ports.

**Condition** is a predicate that can be evaluated at runtime to determine if an operator should execute.

**Executor** manages execution of a Fragment on a physical node.

# X86 Support

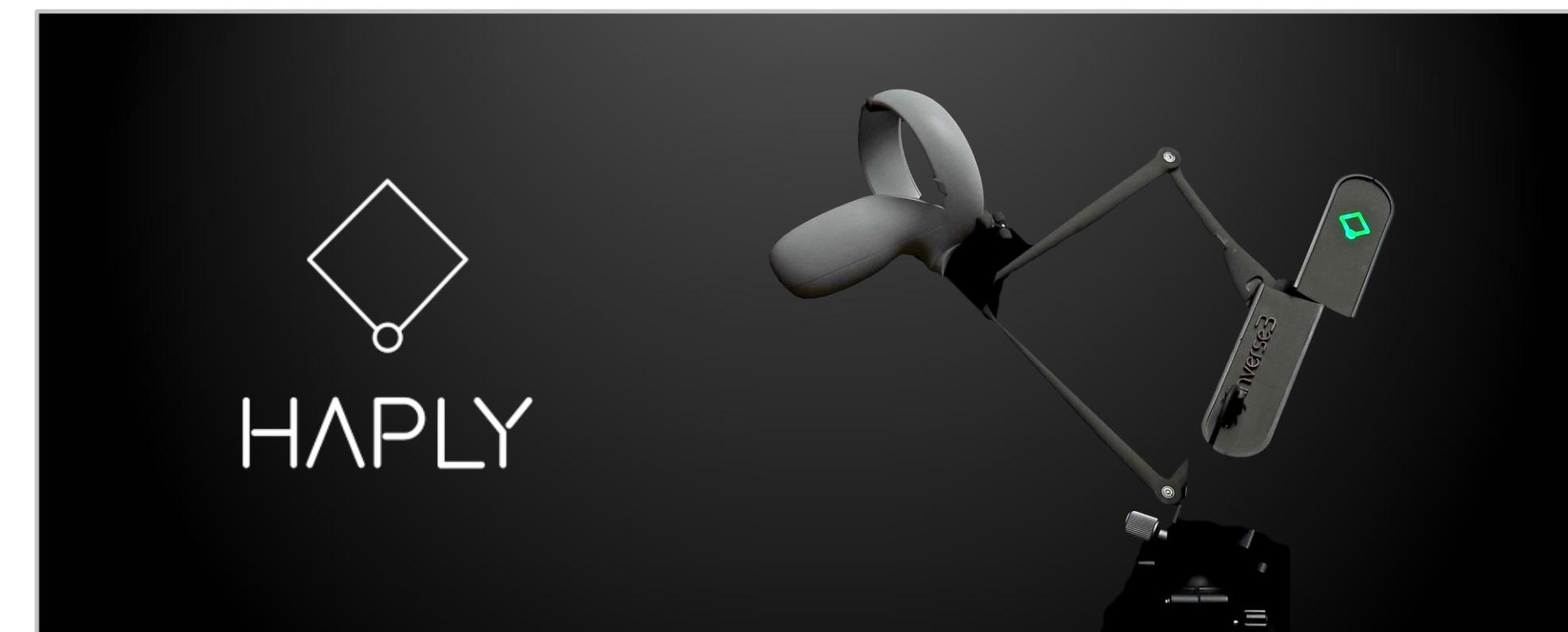
Quickly build AI applications for medical devices on any machine

---



- Ubuntu 20.04 Laptop or Workstation
- Turing or Ampere GPU

# Medical Devices Developing on Clara Holoscan



# Announcing Next Generation Surgery Platforms

Clara Holoscan on NVIDIA IGX selected to deliver proven, real-time AI at the clinical edge



**Activ Surgical**

ActivEdge™

Intelligent light for  
real-time AI and augmented reality



**Moon Surgical**

Maestro Robot

Surgical assistant with  
sensory and perception

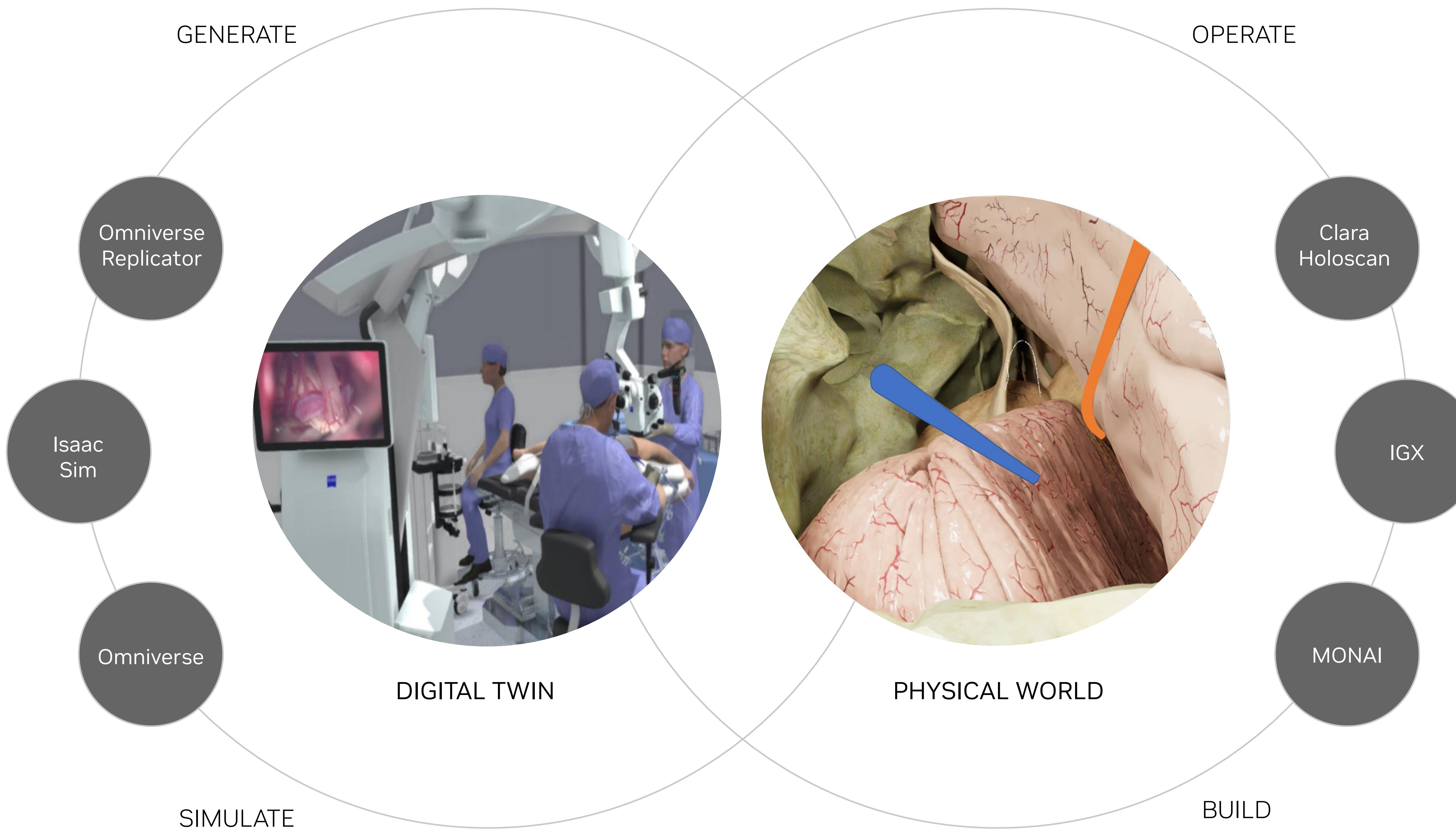


**Proximie**

Operating room telepresence  
for real-time, remote collaboration

# Digital Twins for Surgical Robots

Clara Holoscan integration with NVIDIA Omniverse simulate clinical environments



# Learn More about Clara Holoscan

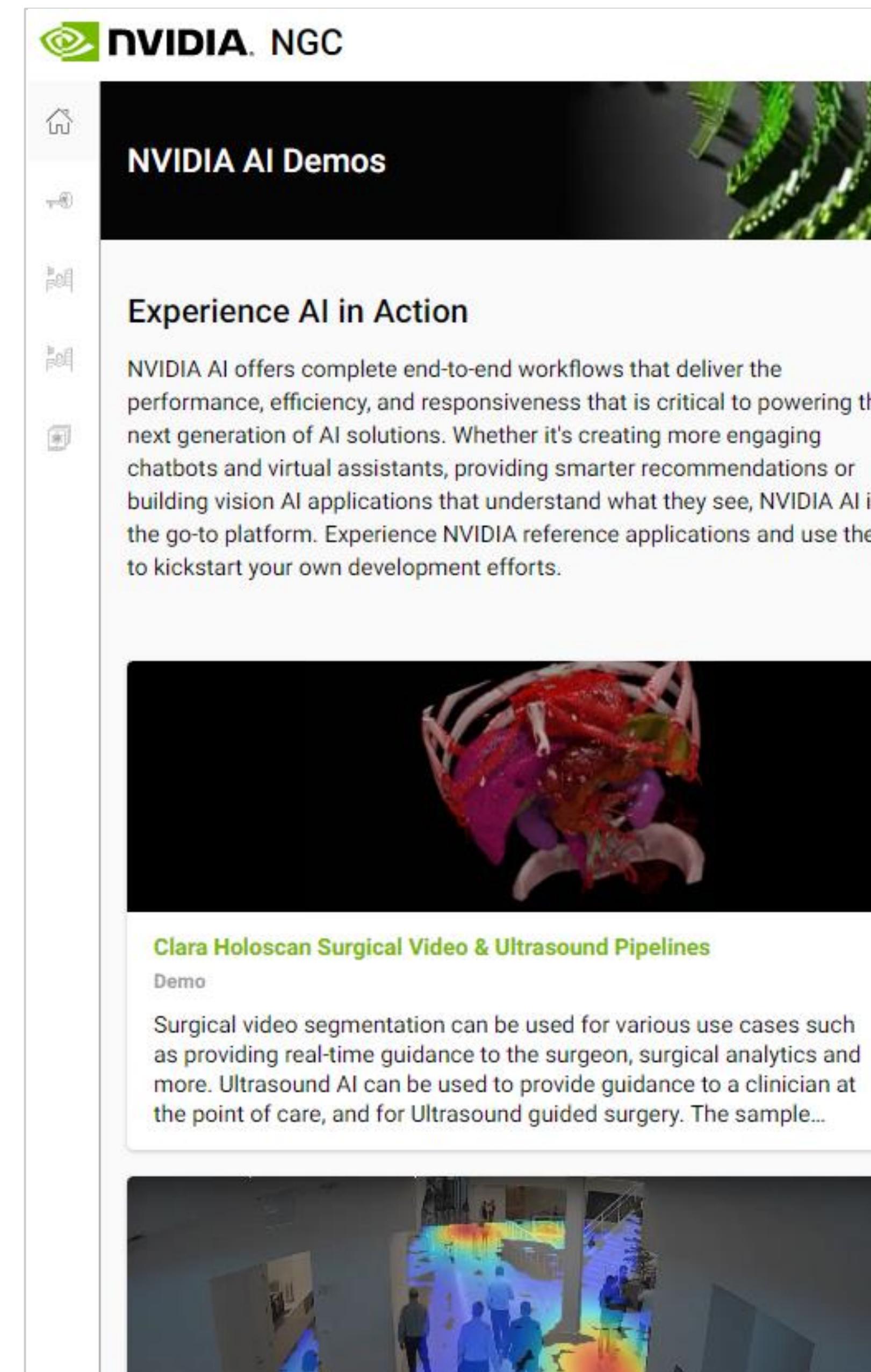


The AI Computing Platform for Medical Devices

Accelerate the next generation of AI-enabled device development with NVIDIA Clara™ Holoscan. A domain-specific AI computing platform, Clara Holoscan delivers the full-stack infrastructure needed for scalable, software-defined, real-time processing of streaming data at the edge—so developers can build devices and deploy AI applications directly into clinical settings.

## Clara Holoscan Webpage

<https://www.nvidia.com/en-us/clara/medical-devices/>



NVIDIA AI Demos

Experience AI in Action

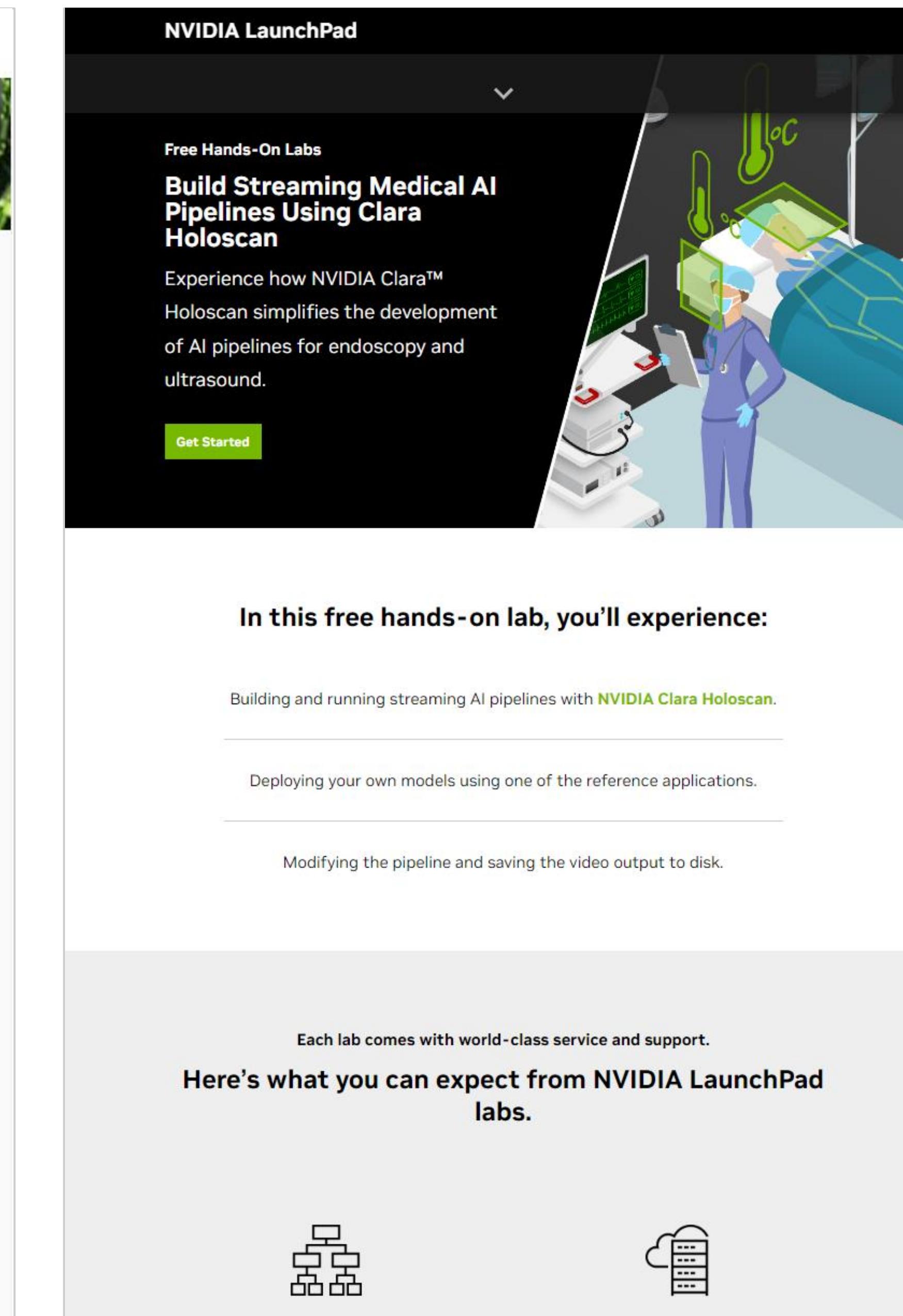
NVIDIA AI offers complete end-to-end workflows that deliver the performance, efficiency, and responsiveness that is critical to powering the next generation of AI solutions. Whether it's creating more engaging chatbots and virtual assistants, providing smarter recommendations or building vision AI applications that understand what they see, NVIDIA AI is the go-to platform. Experience NVIDIA reference applications and use them to kickstart your own development efforts.

Clara Holoscan Surgical Video & Ultrasound Pipelines Demo

Surgical video segmentation can be used for various use cases such as providing real-time guidance to the surgeon, surgical analytics and more. Ultrasound AI can be used to provide guidance to a clinician at the point of care, and for Ultrasound guided surgery. The sample...

## NGC Demo

<https://demos.ngc.nvidia.com/holoscan>



NVIDIA LaunchPad

Free Hands-On Labs

Build Streaming Medical AI Pipelines Using Clara Holoscan

Experience how NVIDIA Clara™ Holoscan simplifies the development of AI pipelines for endoscopy and ultrasound.

In this free hands-on lab, you'll experience:

- Building and running streaming AI pipelines with NVIDIA Clara Holoscan.
- Deploying your own models using one of the reference applications.
- Modifying the pipeline and saving the video output to disk.

Each lab comes with world-class service and support.

Here's what you can expect from NVIDIA LaunchPad labs.



## NVIDIA LaunchPad

<https://www.nvidia.com/try-clara-holoscan/>



OUR MISSION IS:  
POWERING THE WORLD'S MOST IMPORTANT DEVICES®

@Dedicated

MARKETS WE SERVE WHAT WE DO ABOUT US DEDICATED DIFFERENCE

Webinar with Dedicated Computing & NVIDIA

Heterogeneous Computing for AI Enablement within Healthcare and Industrial Applications using NVIDIA IGX

Tuesday October 4th, 2022 12:00 CDT

## Dedicated Computing Webinar

<https://info.dedicatedcomputing.com/dc-gtc-webinar-registration>

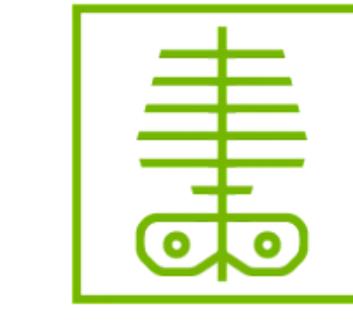
# RESOURCES



NVIDIA Clara Holoscan Platform  
[nvidia.com/en-us/clara/medical-devices/](https://nvidia.com/en-us/clara/medical-devices/)



NVIDIA Clara Holoscan SDK  
[developer.nvidia.com/clara-holoscan-sdk](https://developer.nvidia.com/clara-holoscan-sdk)



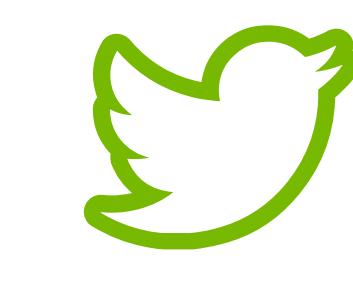
NVIDIA Clara Holoscan on GitHub  
[github.com/NVIDIA/clara-Holoscan](https://github.com/NVIDIA/clara-Holoscan)



NVIDIA Clara DevKit  
<https://developer.nvidia.com/clara-devkit-distributors>



MONAI YouTube Channel  
[youtube.com/c/ProjectMONAI](https://youtube.com/c/ProjectMONAI)



Follow us on Twitter  
[@NvidiaHealth](https://twitter.com/NvidiaHealth)